

INFORMAATIOSTA TIEDON UUOKSI

**Sanna Kumpulainen
ja Mika Sihvonen**

(toim.)

INFORMAATIOSTA TIEDON UUOKSI

**INFORMAATIO-
TUTKIMUKSEN
ILMIÖITÄ
JA UOMIA**

**Vastapaino
Tampere 2025**



Tämä teos on lisensoitu Creative Commons Nimeä-EiKaupallinen-JaaSamoin 4.0 Kansainvälinen -käyttöluvalla. Tarkastele käyttölopaa osoitteessa:
<https://www.creativecommons.org/licenses/by-nc-sa/4.0/legalcode.fi>

Tämän kirjan avoimen julkaisemisen (Open Access) on rahoittanut Tampereen Yliopisto.

DOI: 10.58181/VP9789523972834 (Open Access)
<http://doi.org/10.58181/VP9789523972834>

Maksullinen painettu kirja ja sähkökirja (ePub) ovat myös saatavilla.



VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

© Vastapaino, tekijät

ISBN: 978-952-397-283-4 (painettu kirja)
ISBN: 978-952-397-234-6 (Open access -kirja)

Kustannusosakeyhtiö Vastapaino
Yliopistonkatu 60 A
33100 Tampere
www.vastapaino.fi

Painopaikka: Drukätava 2025

Sisällys

	Aluksi	7
	Kirjoittajat	9
1	Mitä informaatiotutkimus on?	
1.1	Informaatio, ihminen, järjestelmä ja yhteisö	13
1.2	Informaation käsitteen eri merkitykset	25
2	Ihminen ja informaatio	
2.1	Informaatiokäytännöt	47
2.2	Lukutaitojen ja lukemisen tutkimus	63
2.3	Tehtävälähtöinen informaatiovuorovaikutus	87
3	Informaatio ja tiedonhakuja järjestelmät	
3.1	Hakukoneista kielimalleihin	115
3.2	Tiedonhaun tutkimus	135
3.3	Algoritmit ja tiedon löydettävyys	153
4	Informaatio yhteisöissä	
4.1	Tieteellinen julkaiseminen – mitä, miksi ja miten?	183
4.2	Sähköinen asiakirjahallinta hakee vielä uomiaan	211
4.3	Oikeustieteen ja informaatiotutkimuksen yhteyksiä etsimässä	237

Aluksi

Tämä teos sai alkusysäyksensä Tampereen yliopiston informaatiotutkimuksen opetuskokouksessa syksyllä 2021. Opetusväen pohdintoissa totesimme, että vaikka informaatiotutkimuksen alalla tieteellinen julkaisutoiminta onkin monipuolista ja aktiivista, olisi paikallaan kartoittaa erityisesti suomenkielisen kirjallisuuden kanalta tarpeellisia ja ajankohtaisia informaatiotutkimuksen teemoja.

Lähihistoriasta tunnetaan kaksi informaatiotutkimuksen alaa laajasti käsittelevää teosta, jotka ovat työryhmällemme varsin tuttuja: Ilkka Mäkisen vuonna 1999 toimittama *Tiedon tie: johdatus informaatiotutkimukseen* sekä *Ote informaatiosta: johdatus informaatiotutkimukseen ja interaktiiviseen mediaan*, joka ilmestyi Sami Serolan toimittamana vuonna 2010. Nämä aiemmat julkaisut ovat monessa suhteessa edelleenkin mainiota lähdemateriaalia alueen tutkimukseen, ja käsillä oleva teos tuo tässä jatkumossa oman lisänsä kotimaiseen informaatiotutkimuksen kirjallisuuteen.

Siinä missä digitaalinen ja verkkovälitteinen informaatio vielä 1990-luvun puolivälissä edusti satunnaista television, radion, kirjojen, lehtien ja paperilomakkeiden välittämän informaation oheisresurssia, elämme nykyään maailmassa, jossa informaatio on ympärillämme mitä erilaisimmilla elämänalueilla ja mitä erilaisimmissa välineissä. Toimintamme tuottaa, muokkaa ja organisoii informaatiota digitaalisessa maailmassa suoraan tai välillisesti, ja informaation hyödyntäminen sekä arviointi edellyttää meiltä jokaiselta ymmärrystä ja osaamista.

Tämä teos pyrkii tarjoamaan lisävalaistusta sellaisiin informaatiotutkimuksen teemoihin, joista ajankohtainen suomenkielinen kirjallisuus on työryhmämme mielestä tarpeen. Esimerkiksi lukutaitoihin, algoritmeihin, Euroopan unionin tietosuoja-asetukseen ja ChatGPT-tekoälysovellukseen liittyvä viimeaikainen yhteis-

kunnallinen keskustelu ja ilmeinen tiedontarve ovat vaikuttaneet näiden aihepiirien valintaan. Kaikkia teoksen artikkeleita yhdistää pyrkimys tarkastella informaatiota ilmiönä. Artikkelit muodostavat informaation kaaren ja jatkumon, alkaen käsitteenmäärittelyistä ja ihmiseen liittyvistä näkökulmista ja jatkuen aina järjestelmiin sekä yhteisökyksymyksiin asti.

Teoksen kohderyhmää hahmotellessamme mielessämme olivat erityisesti korkeakoulujen opiskelijat, jotka voivat hyödyntää artikkelikokoelmaa eri opintojaksoilla, samoin kuin artikkeleissa esiintyvää kirjallisuutta opinnäytteidensä lähdeaineistona. Samalla teos on suunnattu myös informaatioalan ammattilaisille ja kaikille ajankohtaisista informaatiotutkimuksen teemoista kiinnostuneille.

Teoksen toimittajat Tampereella, Nokia Arenan kupeessa huh-
tikuussa 2025.

Kirjoittajat

Paavo Arvola

FT Paavo Arvola on informaatiotutkimuksen yliopistonlehtori, alanaan digitaaliset ihmistieteet, suuret kielimallit ja tiedonhaku. Hän on kiinnostunut erityisesti tiedonhaun ja tiedonhaun tutkimuksen sovellusalueista ja soveltamisesta. Arvola toimii tutkinto-ohjelmavastaavana Tampereen yliopiston informaatioteknologian ja viestinnän tiedekunnan viestinnän monitieteisessä tutkinto-ohjelmassa.

ORCID: 0000-0001-7209-8033

J. Tuomas Harviainen

FT, MBA J. Tuomas Harviainen toimii informaatiotutkimuksen ja interaktiivisen median professorina Tampereen yliopistossa. Hänen erityisalaansa ovat tietokäytännöt, varsinkin tiedon jakaminen eri konteksteissa.

ORCID: 0000-0002-6085-5663

Pekka Henttonen

YTT, dosentti Pekka Henttonen toimii asiakirjahallinnan yliopistonlehtorina Tampereen yliopistossa. Muun muassa aineistojen siirtyminen digitaaliseen elinkaareen on yksi hänen tutkimusalueistaan.

ORCID: 0000-0003-2600-0378

Heikki Keskustalo

FT Heikki Keskustalo toimii yliopistotutkijana Tampereen yliopistossa. Hänen tutkimusalaansa on tiedonhaku, erityisesti tekstitiedonhaku sekä tehtäväperustainen informaatiovuorovaikutus.

ORCID: 0000-0003-1415-373X

Sanna Kumpulainen

FT Sanna Kumpulainen on informaatiotutkimuksen professori Tampereen yliopistossa. Hänen tutkimuskohteitaan ovat ihmisten ja erilaisten tietoresurssien informaatiovuorovaikutus, tiedonhaku ja hankinta sekä erilaisten avointen tutkimusaineistojen hyödyntäminen. Hän opettaa informaatiotutkimuksen maisteriohjelmassa muun muassa informaatiovuorovaikutusta ja tehtävälähtöisen tiedonhaun teemoja sekä avoimen tieteen teemoja.

ORCID: 0000-0002-7016-257X

Elina Late

FT Elina Late työskentelee Tampereen yliopistossa informaatioteknologian ja viestinnän tiedekunnassa yliopistotutkijana. Hänen tutkimusalueisiinsa kuuluvat tutkimusaineistojen avoimuus ja niihin liittyvät tietokäytännöt sekä tieteellinen viestintä.

ORCID: 0000-0002-3232-1365

Mika Sihvonen

FL Mika Sihvonen toimii Tampereen yliopistossa informaatiotutkimuksen yliopisto-opettajana ja opintosuuntavastaavana. Hän on kiinnostunut yliopisto-opiskelijoiden tavoista hyödyntää eri tietoresursseja. Myös opiskelun ja työelämän edellyttämät aikuisten perustaidot sekä aktiivisen kansalaisuuden teemat ovat olleet hänen tutkimuskohteitaan.

ORCID: 0000-0001-6960-5890

Sanna Talja

YTT Sanna Talja on tiedon organisoinnin ja tiedonhallinnan yliopistonlehtori Tampereen yliopistossa. Hänen tutkimusalueitaan ovat tieteen- ja teknologiantutkimus, informaatiofilosofia sekä tieteellinen kommunikatio. Hänen keskeisiin opetusalueisiinsa kuuluvat verkkopalvelujen informaatioarkkitehtuurin suunnittelu sekä lukemisen tutkimus.

ORCID: 0000-0002-2854-0784

Tomi Voutilainen

Hallinto-oikeuden dosentti, HTT Tomi Voutilainen toimii julkisoikeuden professorina Itä-Suomen yliopistossa. Hänen tutkimus- ja opetusalueinaan ovat informaatio-oikeus, finanssihallinto-oikeus ja hallinto-oikeus.

Mitä informaatio- tutkimus on?

1.1 Informaatio, ihminen, järjestelmä ja yhteisö

Sanna Kumpulainen

- Informaatiotutkimus tarkastelee tallennetun tiedon käyttöä, sen organisointia ja sen tuottamia hyötyjä ihmisille ja yhteisöille.
- Historiallisesti informaatiotutkimus on syntynyt kahden eri tutkimushaaran, kirjastotieteen ja dokumentaation, yhdistymisestä.
- Informaation virtoja voidaan tarkastella järjestelminä.
- Informaatiotutkimus on monialainen ja monimuotoinen tieteenala, jossa tarkastelun keskiössä ovat informaatio ja sen virrat.

Jo muutama vuosikymmen sitten informaation todettiin tulleen hengittämämme ilman kaltaiseksi – se ympäröi meidät meidän sitä huomaamatta, emmekä voi elää ilman sitä.¹ Tuolloin tuskin kyettiin edes kuvittelemaan nykyisen kaltaista informaatioympäristöä, jossa arkinen sosiaalinen ympäristö sulauttaa virtuaalisen maailman fyysiseen maailmaan saumattomasti. Suomalaisen yhteiskunnan palvelut ovat siirtyneet digitaaliseen ulottuvuuteen, ja tietotyötä on mahdollista tehdä pääosin verkossa. Tässä digitaalisessa maailmassa informaatio virtaa eri palveluiden välillä, mutta loppujen lopuksi kaiken tämän informaation tarkoitus on palvella ihmisiä ja inhimillistä toimintaa. Tekoälyt lisäävät oman suolansa sekä oman toimijuutensa tähän verkostoon. Töitä ulkoistetaan nyt jo erilaisille algoritmisille toimijoille muun muassa päätöksenteon tueksi. Oikean informaation saaminen korostuu ja hankaloituu informaatio-

1 Tague-Sutcliffe 1995.

tulvan ja tekoälyn generoiman informaation vuoksi. Sujuva pääsy oikeaan informaatioon on keskeistä onnistuneelle toiminnalle, ainakin tiedon tarvitsijan ja etsijän näkökulmasta.

Informaatiotutkimuksen tavoitteena on kehittää käsitteitä, menetelmiä ja välineitä, joiden avulla informaatio voidaan saattaa vaittomasti sitä tarvitseville. Informaatiotutkimus tarkastelee toisaalta ihmisten tiedonhankintaa ja -käyttöä erilaisissa tilanteissa ja ympäristöissä ja toisaalta sitä, kuinka informaatiota ja tietoa organisoidaan, hallitaan ja säilytetään. Informaatiotutkimus on luonteeltaan monitieteinen lainaillen piirteitä useilta eri aloilta, kuten tietojenkäsittelytieteistä, yhteiskuntatieteistä, psykologiasta ja organisaatiotutkimuksesta, mutta samalla keskittyen informaatioon ja sen rooliin yhteiskunnassa.

Informaatiotutkimus on vahvasti yhteiskuntatieteellinen ala, mutta eri tutkimushaarojen välillä on suuria eroja. Yhteisinä tavoitteina kuitenkin on tarkastella ja löytää keinoja tukea ja ymmärtää informaation tuottamista, jakamista ja löytämistä sekä säilytystä ja järjestämistä. Tämä mahdollistaa inhimillisen toiminnan, tietämyksen muodostamisen sekä päätöksenteon auttamisen ja ”parantamisen” erilaisissa ympäristöissä ja tilanteissa. Informaatiotutkimus tarkastelee informaatioilmiöitä laajasti, myös kriittisestä, yhteiskunnallisesta sekä tieteenfilosofisesta näkökulmasta.

Informaatiotutkimus on vaikuttanut myös muihin tieteenaloihin. Tiedonhaun evaluointimittareilla on ollut varsin valtava vaikutus siihen, miten esimerkiksi tietojenkäsittelytieteissä tutkitaan algoritmien toimintaa tai neuroverkkojen tuottamia tietopalasia. Informaatiotutkimuksesta peräisin olevia mittareita ovat muun muassa kumuloitu hyöty -mittaristot,² perinteinen tiedonhaun tutkimusparadigma,³ relevanssin käsite⁴ sekä tehtävälähtöiset tutkimukset ja tutkimusasetelmat,⁵ joita käsitellään tarkemmin luvuissa

2 Järvelin & Kekäläinen 2002.

3 Cleverdon 1970.

4 Hjørland 2010; Saracevic 1975.

5 Ks. Vakkari 2003.

2.3, 3.1 ja 3.2. Lisäksi muun muassa organisaatio- ja hallintotieteiden, terveystieteiden, mediatutkimuksen ja journalistiikan puolella ovat vaikuttaneet esimerkiksi tiedonlähteiden, tiedontarpeiden ja käytön eli laajemmin informaatiokäyttämisen ja -käytäntöjen tutkimus⁶ (tietokäytännöistä lisää luvussa 2.1), joka kuuluu informaatiotutkimuksen ydinalueisiin. Tieteellisen viestinnän tutkimus on läpäissyt luonnollisesti kaikki tutkimusalat, sillä esimerkiksi bibliometriikka⁷ hyödynnetään kaikilla tutkimuksen aloilla (tieteellisestä viestinnästä luvussa 4.1). Bibliometriikan tausta on tietokantojen indeksoinnissa ja kirjastojen kokoelmien evaluoinnissa.

Informaatiotutkimuksen taustaa

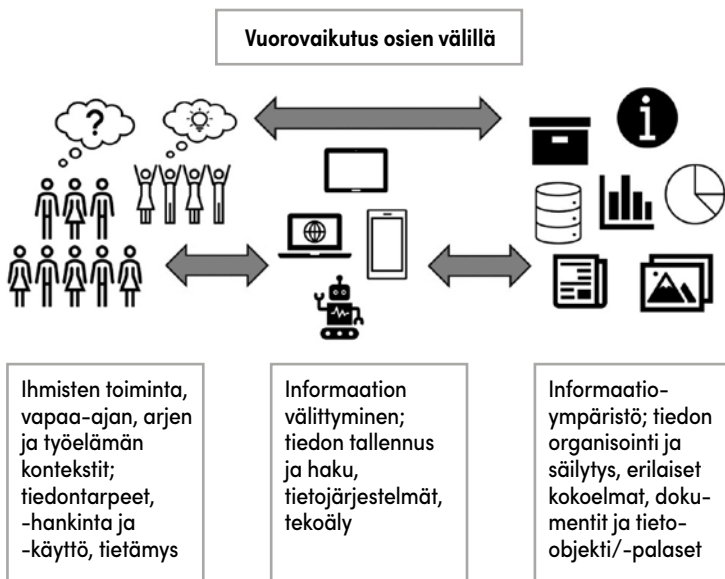
Historiallisesti informaatiotutkimus on syntynyt kahden eri tutkimushaaran eli kirjastotieteen ja dokumentaation yhdistymisestä. Kirjastotieteen synty pohjautui tarpeeseen järjestää kasvavia kirjamääriä ja kouluttaa osajia tähän tarkoitukseen. Käytännön toimintojen kehittäminen muuttui tieteelliseksi tutkimukseksi 1800-luvun lopulla, jolloin Göttingenin yliopistoon perustettiin alalle professuuri. Tavoitteena oli kouluttaa kirjastonhoitajia tieteellisiin kirjastoihin. Dokumentaatio taas syntyi pyrkimyksestä analysoimaan, tallentamaan ja hakemaan tietoa sen kaikissa muodoissa sekä saattamaan tietoa aktiivisesti käyttäjille. Dokumentaatiosta kehittyi tietopalvelu ja informatiikka ja kirjastotieteen ja informatiikan yhdistelmästä nykyinen informaatiotutkimus.⁸

Informaatiotutkimuksessa ollaan kiinnostuneita ihmisten informaatiotoiminnasta, erilaisista tiedonhaku- ja tallennusjärjestelmistä sekä toiminnan informaatioympäristöistä, jotka koostuvat erilaisista informaatiokokoelmista, tietokannoista ja informaatiopalasista. Kuvio 1 kuvaa tätä moninaista ilmiökenttää ja esittää yhdenlaisen tavan hahmottaa informaatiotutkimuksen tarkastelukohteita. Ku-

6 Ks. Choo 1996; Savolainen 1995.

7 Garfield 1955.

8 Mäkinen 2010.



Kuvio 1. Informaatiotutkimuksen tutkimuskohteita.

viossa on pyritty kattamaan kokonaisuus, joten se jää pakostakin tietyiltä osin varsin yleiseksi. Tavoitteena on kattaa kirjan kohteena olevan ilmiökentän osasia ja niiden välisiä suhteita.

Informaatiotutkimuksessa tarkastellaan kuviossa esitettyjen eri osa-alueiden välistä vuorovaikutusta eli esimerkiksi erilaisten ihmisryhmien tiedonhankintaa ja tiedontarpeita suhteessa tarjolla olevaan informaatioon, erilaisten tiedonhakujärjestelmien käyttöä tai sitä, miten erilaiset kokoelmat tulisi järjestää, jotta niiden informaationvälitysohjelmat olisivat parhaiten ihmisten löydettävissä. Lisäksi nämä osa-alueet itsessään ovat mielenkiinnon kohteina. Silloin tarkastellaan esimerkiksi sitä, millaisia erilaisia määrittäviä tai toiminnallisia tekijöitä liittyy kuhunkin kohteeseen. Näitä voivat olla inhimillisessä kontekstissa muun muassa yhteiskunnan muutokset ja niiden synnyttämät tiedontarpeet, työtehtävien ominaisuudet ja niiden vaikutus tiedonhankintaan tai erilaisten ryhmien informaatio-

tiokäytäntöjen piirteet. Tietoa välittävien järjestelmien osalta kiinnostavia ovat esimerkiksi järjestelmän piirteet ja kyky listata mielekkäitä tuloksia tai tekoälyn avulla generoitujen sisältöjen laatu suhteessa tiedonhakujärjestelmään. Informaatioympäristöä tutkittaessa tarkastelukohteita voisivat puolestaan olla esimerkiksi yhden tai useamman tietoresurssin organisointi (metadata ja luokitukset), organisoinnin konteksti (yhteiskunnalliset/organisaation kontekstit) tai kokoelman hallinta ja kehittäminen.

Informaatio ja järjestelmät

Kutakin osa-aluetta – ihmisten toimintaa, informaation välittymistä sekä informaatioympäristöä – voidaan tarkastella informaatiojärjestelmänä.⁹ Tämä koskee myös kaikkien osa-alueiden muodostamaa kokonaisuutta. Systeemiajattelussa yksi tarkastelujako voi olla palvelu, jolloin keskeistä on ottaa huomioon palveltavan järjestelmän (esimerkiksi ihmisten toiminta jossain kontekstissa tai tietyn organisaation toiminta) piirteet, jotta voidaan suunnitella ja muokata palvelevan järjestelmän (kirjasto, arkisto, tietopalvelu tai tiedonhakujärjestelmä) ominaisuuksia vastaamaan näihin tarpeisiin.¹⁰ Tässä kohtaa olisi hyvä huomioda, että kirjastot ja arkistot ovat myös informaatiojärjestelmiä, vaikka nykykielessä tietojärjestelmällä tarkoitetaan usein pelkästään jonkinlaista tietoteknistä järjestelmää. Systeeminäkökulma sitoo tämän kirjan teemat saman ilmiökentän osiksi.

Informaatio virtaa tässä kokonaisuudessa osana järjestelmää tai järjestelmiä moniin suuntiin – ihmisiltä tiedon tuottajina erilaisiin tietokokoelmiin, kokoelmien ja tietojärjestelmien välillä, jolloin tietoja yhdistellään joko automaattisesti tai ihmisten toiminnan tulokseksi uudeksi tiedoksi. Lisäksi informaatiota tietenkin etsitään erilaisista kokoelmista ihmisten, yhteisöjen ja yhteiskunnan käyttöön.

9 Buckland 1991; Checkland & Holwell 1997; Vakkari 1999.

10 Vakkari 1999.

Informaatioympäristöjen monimuotoistumisen, jokapaikkaistumisen ja samanaikaistumisen (ubiikkisuus) sekä informaation kaikkialle tunkevuuden (pervasiivisuus) myötä informaatiota on tarjolla hyvinkin paljon. Esimerkiksi monet tekevät hakuja jopa huomaamattaan ja suositteualgoritmit tuottavat räätälöityjä tarjoumia informaatioon ilman vaivannäköä.¹¹ Usein kuvia tai linkkejä klikkailemalla käynnistyy erilaisia suodatus- ja suositteelumekanismeja.

Sekä tarkoituksellinen että tahaton tiedonhaku aiheuttaa informaatiotulvaa ja siitä seuraavaa informaatioähkä, mutta siitä saatetaan seurata myös digitaalisia kuiluja erilaisten ryhmien mahdollisuuksien välille digitaalisen osallisuuden näkökulmasta. Toisilla on parempi kyky ja mahdollisuudet käsitellä ja hallita digitaalista informaatiota kuin toisilla. Tämä näkyy, jos esimerkiksi vertaillaan akateemista työyhteisöä, jolla on hyvät digitaaliset palvelut käytössään sekä taidot niiden hyödyntämiseen, ja laitoshoidossa olevia vanhuksia, joiden tietotekniset kyvyt saattavat olla puutteellisia tai epäajantasaisia mutta joilta kuitenkin odotetaan digiyhteiskunnan palveluiden käyttöä oman elämän asioiden hoitamiseen.

Järjestelmien sisällä virtaava informaatio itsessään on varsin moniselitteinen käsite ja asettaa haasteita sen tutkimiselle. Tämän kirjan luvussa 1.2 avataan informaation käsitettä sekä teoreettisia keskusteluita informaatiosta ja sen tulkinnoista. Informaatio voidaan ottaa haltuun määrittelemällä, mitä sillä kulloisessakin tarkastelukulmassa ja -asetelmassa tarkoitetaan ja onko jotain mahdollisia ulottuvuuksia rajattu pois.

Yksi esimerkki järjestelmä- tai systeemiajattelusta voisi olla, että jos tunnistetaan tietyn ryhmän johonkin toimintaan tarvitsemia tiedonlähteitä, voidaan arvioida sitä, miten hyvin yksittäinen tiedonlähde tai kanava vastaa toiminnasta kumpuavaan tiedontarpeeseen. Tämä voi tarkoittaa myös sitä, että jokin muu tapa hankkia tai tarjota tietoa voisi olla parempi kuin se, mikä oli ryhmän käytössä tässä tarkastelussa. Tällä on seurauksia tietopalveluiden ja tiedonlähteiden

11 Ks. lisää esim. Haider & Sundin 2019.

den tarjoamiselle ja suunnittelulle. Tiedonlähteiden ja kanavien tarkastelu onkin keskeistä esimerkiksi tietokäytäntöjen ja tiedonhankinnan tutkimuksessa. Informaatiokäytäntöjä tarkasteltaessa systeeminä eli järjestelmänä on ihmisten arjen järjestelmä. Siinä syntyy tiedontarpeita, joihin hankitaan ja luodaan informaatiota erilaisista lähteistä, kuten netin yhteisöistä. Informaatiokäytäntöjä tarkastellaan luvussa 2.1.

Monenlaiset lukutaidot ovat keskeisiä kansalaistaitoja monimuotoisessa digitaalisessa maailmassa. Informaation ”siirtyminen” ihmiseen erilaisista informaatioympäristöistä ei ole suoraviivainen prosessi, vaan edellyttää ymmärryksen muodostamista erilaisista välittymismekanismeista, informaatioympäristöstä ja informaation luonteesta.¹² Ymmärryksen muodostamista voidaan lähestyä lukutaitojen käsitteen avulla, jossa lukeminen tarkoittaa muutakin kuin puhtaasti tekstitaitoja. Lukutaidot edellyttävät informaation hankkijalta ja hyödyntäjältä muun muassa tekstien auktoriteettien ymmärtämistä ja ymmärryksen dialogista rakentumista suhteessa tiedonlähteeseen. Tässä vuorovaikutuksessa siihen, mikä on ymmärryksen rakentumisen lopputulos, vaikuttavat sekä tiedon hankkijan eli lukijan taustat että informaatiolähteen taustat ja yksilön ymmärrys näistä. Tätä monimutkaista ilmiökenttää eli lukemista ja lukemisen tutkimusta tarkastellaan luvussa 2.2.

Ihmisten elämänpiiriin kuuluu myös usein työ, ja yksi informaatiotutkimuksen pitkäaikaisista kiinnostuksen kohteista on ollut monenlaisen tietotyön tukeminen informaation avulla. Työyhteisöt muodostavat kukin oman maailmansa, jossa omaksutut kulttuurit, käytännöt ja säännöt vaikuttavat toimintaan. Työyhteisön kuvaaminen järjestelmänä tarkoittaa sen tarkastelua kokonaisuutena, joka koostuu useista vuorovaikutuksessa olevista osista. Tämä auttaa ymmärtämään, miten eri komponentit ja toiminnot liittyvät toisiinsa ja vaikuttavat koko yhteisön toimintaan. Luku 2.3 käsittelee tehtävälähtöistä informaatiovuorovaikutusta, jossa kuvataan si-

12 Dervin 1983.

säkkäisiä ja rinnakkaisia informaatioaktiviteetteja. Tehtävä, joka on ollut tätä mallia luodessa mallin tekijöillä mielessä, oli työympäristöstä poiketen oppimistehtävä. Tietointensiivisissä työtehtävissä on kuitenkin monia oppimistehtävien piirteitä, joten malli sopii myös monenlaisten työtehtävien jäsentämiseen. Tässä lähestymistavassa on keskeistä ajatus tehtävän prosessien ja toimintojen ymmärtämisestä ja niiden tukemisesta informaatiolla ja tietoresursseilla. Malli ei lähde liikkeelle organisaatiosta eikä työpaikan tarpeista, vaan yksilön toiminnasta sekä tarpeista ja näiden mallintamisesta. Tehtävän rajaamassa järjestelmässä tieto virtaa tietoresursseista, sekä sisäisistä että ulkoisista, monenlaisiin aliprosesseihin ja sieltä työntekijän käyttöön ja koko työpaikan ja yhteisön hyödyksi. Tieto on jossain, josta se haetaan, valitaan ja hyödynnetään jonkinlaisen tuotoksen aikaansaamiseksi. Tällä tuotoksella on jokin perimmäinen tarkoitus joko yksilölle itselleen, organisaatiolle tai yhteisölle.

Informaatiotutkimuksessa on nykyisin vaikea välttyä teknologioilta digitalisoitumisen vuoksi, ja yksi vahva tutkimustraditio onkin tiedonhaun tutkimus. Se kattaa niin tiedonhakujärjestelmät kokonaisuutena kuin niiden osaset, hakujen tekemisen, kokoelmat, informaationsiällöt ja dokumentit, indeksit sekä täsmäytys-, sijajärjestys- ja suositte-lujärjestelmät. Tässä tietoteknisillä ratkaisuilla on suuri rooli. Tämä ilmiökenttä on se, jota tarkoitetaan, kun puhekielessä puhutaan ”informaatiojärjestelmistä”. Tiedonhaun tutkimuksessa tarkasteltava ilmiö rajataan yleensä dokumenteista ihmisen järjestelmiin tehtyihin syötteisiin eli kyselyihin, jolloin ihmisen toiminta hakijana ja toimintaympäristö saatetaan rajata pois, samoin dokumenttien tuottajat ja kokoelmien luominen sekä siihen liittyvät toimintaympäristöt. Kyselyiden ja dokumenttien välistä suhdetta mitataan erilaisilla mittareilla, jotka mittaavat hakujärjestelmän ominaisuuksien ”hyvyyttä” esimerkiksi suhteessa muihin järjestelmiin. Nykyisin suuret kielimallit ja niiden keskustelukäyttöliittymät tuottavat myös niin sanotusti valmiiksi pureskeltua tietoa eli synteetin omaista tulostetta tietovarannoistaan tiedonhakujärjestelmien

perinteisen järjestetyn listan sijasta. Tätä ilmiökenttää avataan luvuissa 3.1 ja 3.2.

Laajentamalla tarkastelukulmaa tietojärjestelmien ja informaatioympäristön ominaisuuksien ja kontekstien vaikutuksista niiden käyttäjyhteisöihin saadaan uudenlainen sosiotekninen näkökulma, jossa tietojärjestelmiä ei tarkastella vain yksiselitteisinä tutkimuksen kohteina vaan myös kriittisesti yhteiskuntaan vaikuttavina tekijöinä. Tietojärjestelmien suositusalgoritmeja, suodatusta sekä erilaisia järjestelyalgoritmeja ja käyttäjäprofileja suhteessa niiden yhteiskunnallisiin ja käyttäjiin kohdistuviin vaikutuksiin tarkastellaan luvussa 3.3.

Informaatioyhteisöillä on erilaisia omia sääntöjä ja käytäntöjä, jotka määrittävät yhteisön toimintaa sekä informaation merkitystä kussakin yhteisössä ja vaikuttavat siihen, millainen kustakin informaatiojärjestelmästä syntyy. Tiedeyhteisö on yksi erittäin tietointensiivinen kokonaisuus, jonka tarvitsemaa tietoa, tiedonkäyttöä sekä tuottamaa tietoa tutkitaan paljon. Tämän koko ekosysteemin eli tiedejärjestelmän tutkimus sisältää lisäksi tiedeyhteisöjen toiminnan tarkastelun. Erilaiset toimintaa määrittävät reunaehdot ja kehityskaaret ovat olleet informaatiotutkimuksen kiinnostuksen aiheita. Tiedeyhteisöjen ja tieteellisen julkaisemisen tutkimus on ollut keskeistä tieteellisten kirjastojen ja tietojärjestelmien kehittämisessä sekä erilaisten tutkimuksen tukipalveluiden luomisessa. Tieteellisestä julkaisemisesta ja sen tutkimuksesta kerrotaan tarkemmin luvussa 4.1.

Oman tarkastelunäkökulmansa informaatioon ja sen merkitykseen tuo asiakirjahallinta ja sen tutkimus, jota tehdään edelleenkin vain vähän Suomessa. Arkistot, niiden muodostuminen ja muodostajat sekä arkistomateriaalien monenlaiset käyttäjäkunnat muodostavat monimuotoisen informaatiojärjestelmän. Asiakirjoilla on erityispiirteensä todistusvoimaisina dokumentteina, ja niiden hallintaan ja luomiseen liittyy säännöstöjä sekä monenlaisia käytäntöjä. Digitalisaatio tässä ympäristössä ei ole ehkä niin suoraviivaista kuin muualla, mikä johtuu osittain tämän tiedon ja arkistojen erityispiir-

teistä. Sähköistyminen etenee kuitenkin vauhdilla, ja tätä muutosta tarkastellaan luvussa 4.2.

Erityisen informaatioyhteisön muodostaa oikeustiede ja sen tutkimuskohteet. Osa oikeustieteen tutkimuksesta on informaatiotutkimusta. Oikeusinformatiikassa oikeus nähdään informaatiojärjestelmänä, jossa oikeudellista tietoa tarkastellaan tiedon saatavuuden ja haettavuuden sekä oikeustietojärjestelmien toiminnan näkökulmista, osana lainsäädännön prosesseja sekä informaatio-oikeuden näkökulmista. Informaatio-oikeudellinen sääntely koskee kaikkea informaatioon liittyvää toimintaa yhteiskunnassa, yhtenä esimerkkinä eurooppalainen tietosuojasetus. Oikeustieteen, oikeudellisen informaation ja informaatiotutkimuksen rajapintoja käsitellään luvussa 4.3.

Informaatiotutkimus ei ole vain yksi lähestymistapa tai tutkimusperinne, vaan informaation tarjontaan, hankintaan ja löytymiseen on monta näkökulmaa. Informaatiota voidaan tutkia ihmisen, ihmisen sosiaalisen kontekstin ja informaation välisenä suhteena, jolloin ihmisen toiminta on keskiössä. Jos tarkastelukulma vaihdetaan informaatiota välittäviin teknologioihin, tutkimuksen luonne muuttuu ja tutkimuskohteena on se, miten erilaiset tiedonhakujärjestelmät välittävät informaatiota. Kuitenkin kullakin erilaisella yhteisöllä on sosiaalinen toimintaympäristönsä, joka määrittää kyseisen yhteisön toimintaa ja informaatiokäytäntöjä.

Tämä teos pyrkii antamaan kurkistuksen informaatiotutkimuksen eri ulottuvuuksiin ja osaltaan keräämään yhteen erilaisia tutkimusperinteitä. Yhteiskunnan digitalisoituessa erilaiset informaation vuolaat virrat ovat kuitenkin jo yhtenäistäneet informaatiotutkimusta, ja kaikilla osa-alueilla on yhteisiä rajapintoja enemmän kuin erottavia. Erottavana tekijänä näyttäisi olevan lähinnä näkökulmien painotus. Yhteisenä tekijänä on kuitenkin informaatio, sen saattaminen tarjolle ja sen käyttöön saaminen erilaisten ihmisten ja yhteisöjen hyödynnettäväksi.

Lähteet

- Buckland, Michael (1991) *Information and information systems*. USA: Bloomsbury Publishing.
- Checkland, Peter & Sue Holwell (1997) *Information, systems and information systems: Making sense of the field*. New York: Wiley.
- Choo, Chun Wei (1996) The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International journal of information management*, 16:5, 329–340.
- Cleverdon, Cyril (1970) Progress in documentation: Evaluation tests of information retrieval systems. *Journal of documentation*, 26:1, 55–67. <https://doi.org/10.1108/ebo26487>
- Dervin, Brenda (1983) An overview of sense-making research: Concepts, methods, and results to date. *The International Communication Association annual meeting*, Dallas, Texas.
- Garfield, Eugene (1955) Citation indexes for science. *Science*, 122:3159, 108–111. <https://doi.org/10.1126/science.122.3159.108>
- Haider, Jutta & Olof Sundin (2019) *Invisible search and online search engines: The ubiquity of search in everyday life*. United Kingdom: Routledge.
- Hjørland, Birger (2010) The foundation of the concept of relevance. *Journal of the American society for information science and technology*, 61:2, 217–237. <https://doi.org/10.1002/asi.21261>
- Järvelin, Kalervo & Jaana Kekäläinen (2002) Cumulated gain-based evaluation of IR techniques. *ACM transactions on information systems*, 20:4, 422–446. <https://doi.org/10.1145/582415.582418>
- Mäkinen, Ilkka (2010) Informaatiotutkimuksen tie. Teoksessa Sami Serola (toim.) *Ote informaatiosta*. Helsinki: BTJ Kustannus, 11–74.
- Saracevic, Tefko (1975) Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American society for information science*, 26:6, 321–343. <https://doi.org/10.1002/asi.4630260604>
- Savolainen, Reijo (1995) Everyday life information seeking: Approaching information seeking in the context of “way of life.” *Library & information science research*, 17:3, 259–294. [https://doi.org/10.1016/0740-8188\(95\)90048-9](https://doi.org/10.1016/0740-8188(95)90048-9)
- Tague-Sutcliffe, Jean (1995) *Measuring information: An information services perspective*. San Diego: Academic Press.
- Vakkari, Pertti (1999) Tiedonhankinnan tukeminen ja informaatiotutkimus. Teoksessa Ilkka Mäkinen (toim.) *Tiedon tie. Johdatus informaatiotutkimukseen*, 9–31.
- Vakkari, Pertti (2003) Task-based information searching. *Annual review of information science and technology*, 37:1, 413–464. <https://doi.org/10.1002/aris.1440370110>

1.2 Informaation käsitteen eri merkitykset

J. Tuomas Harviainen

- Sana informaatio viittaa eri tutkimustraditioissa eri asioihin, jotka voivat ensi näkemältä vaikuttaa samankaltaisilta, vaikka eivät sitä oikeasti ole.
- Monet määritelmistä tulevat informaatiotutkimuksen ulkopuolelta.
- On usein hyödyllistä käyttää informaatiosta työmääritelmiä, joilla kerrotaan, mitä sanalla kyseisen tekstin kontekstissa tarkoitetaan.

Sanaa *informaatio* käytetään arkikielessä usein määrittelemättä sitä tarkemmin. Tyypillisiä merkityksiä ovat joko a) se, mitä voimme päätellä eleistä, kielestä, teksteistä ja muista asioista, tai b) kommunikaation materiaaliset muodot, kuten bitit, kirjat, paperit ja muut fyysiset tallenteet.¹ Tieteellisessä ja teknisessä käytössä informaatiolla voidaan puolestaan viitata useisiin eri asioihin, jotka muistuttavat toisiaan, mutta nousevat eri tutkimustraditioista. Tämä yhdistelmä hämää herkästi varsinkin akateemista lukijaa, koska näennäisesti yhteensopivat teoriat eivät tosiasiasa puhu samasta asiasta.

Useat tutkijat ovat käsitelleet näitä eroja, tosin voimakkaasti omista taustoistaan lähtien.² Tästä on seurannut se, että kyseiset tutkijat ovat luoneet jaotteluja ja määritelmiä, jotka heijastavat heidän omien ydinalueidensa käsitteistöä ja oletuksia. Buckland menee jopa niin pitkälle, että hän suosittelee käyttämään informaation sijasta aina jotakin muuta, täsmällisempää sanaa kussakin sen eri

¹ Buckland 2017, 6.

² Ks. esim. Bates 2018; Floridi 2011.

tulkinnoista.³ Vastaavasti esimerkiksi Case ja Given katsoivat, ettei informaatiota, niin sanotusti primitiivisenä käsitteenä, tarvitse edes määritellä, vaan ihmiset ymmärtävät selittämättäkin, mitä sillä suunnilleen tarkoitetaan eri konteksteissa.⁴ Myöhemmässä painoksessa he päätyvät toteamaan,⁵ Bosancicin ja Matijevicin katsausartikkeliin viitaten,⁶ että informaatiolle tuskin löytyy määritelmää, joka kattaisi eri tarpeet ja käyttöyhteydet. Belkinin⁷ tavoin hekin suosittelevat siksi tarkan määrittelemisen sijasta käsitteen avointa tulkintaa, joka on ennen kaikkea kattava ja mahdollistaa tutkimuksen tekemistä informaatiokäyttäytymisen eri muotoihin.

Sanan informaatio taustalla on latinan muotoilua ja muotoon asettamista tarkoittava *in-formare*, joka pohjautuu aristoteeliseen metafysiikkaan, jossa asian olemuksen eli muodon havainnointi tekee havainnoijan siitä tietoiseksi.⁸ Tässä luvussa keskityn informaation käsitteen käyttöön informaatio- ja viestintätieteissä sekä niiden lähialueilla. Kiinnostunut lukija löytää lisää informaation merkityksiä esimerkiksi juridiikan, termodynamiikan tai biologian kentiltä, mutta niille kaikille tämä kirja on aivan liian lyhyt.

Informaation käsite eri tutkimusperinteissä

Oleellisesti kysymys tiivistyy siihen, tarkoitetaanko informaatiolla esimerkiksi viestitettävän signaalin oleellista osaa, jotakin universaalia, joka on läsnä kaikessa ja kaikkialla ja joka sitten valikoivasti voidaan havaita, vai kenties jotakin, joka koskee ihmisten ja/tai dokumenttien ja erilaisten tietojärjestelmien välistä vuorovaikutusta. Kasvavissa määrin etuliitteellä *informaatio-* tarkoitetaan käytännössä myös digitaalisuutta. Esimerkki tästä on *informaatioyhteiskunta*, joka

3 Buckland 2017.

4 Case & Given 2016; ks. myös Bates 2018.

5 Given ym. 2023.

6 Bosancic & Matijevic 2019.

7 Belkin 1978.

8 Haasio & Savolainen 2004.

on digitaalisen tiedon läpäisemä.⁹ Vastaavasti puhutaan *informaatiohyödykkeistä* (*information goods*), jotka viittaavat musiikkitiedostojen tai pelien virtuaaliesineiden kaltaisiin asioihin. Ne koostuvat koodatusta informaatiosta ja ovat periaatteessa monistettavissa loputtomasti, mutta niiden saatavuutta usein rajoitetaan keinotekoisesti.¹⁰

Kun informaatiohyödykkeet – tai informaatio laajemminkin – ovat vapaasti saatavilla eivätkä keinotekoisesti rajoitettuina, ne, joilla on niihin nopein pääsy, voivat saavuttaa liiketoiminnallisia etuja, esimerkiksi uusien innovaatioiden muodossa. Tämä koskee erityisesti niin sanottuja tietointensiivisiä teknologian aloja sekä tietojärjestelmien mahdollistamaa alustataloutta. Myös internet ja sosiaalinen media korostavat tätä varhaisen informaation saannin etua.¹¹ Monet informaatiotutkijat työllistyvät nykyisin erilaisiin asiantuntijatehtäviin liiketalouden aloille, jotka hyödyntävät informaatiota. Siksi on hyödyllistä tuntea, mitä käsitteellä siinäkin kontekstissa tarkoitetaan.

Näkökulmaeroista riippumatta voidaan todeta, että informaatio on kaikille informaatiotieteille – ja myös monille muille aloille – kriittisen keskeinen käsite. Se on myös erittäin vaikea käsite, jonka määrittely on todella haastavaa.¹² Lisähankaluuksia tuottaa myös se, että käsitteille ei ole välttämättä suomenkielisiä vastineita ja esimerkiksi sanoja *informaatio* ja *tieto* saatetaan käyttää suomen kielessä toistensa vastineina (esimerkiksi *tietokäytännöt*) sen sijaan, että niistä muodostuisi selkeä hierarkia.¹³ Monet määritelmistä myös nojaavat monimutkaiseen taustakäsitteistöön, joiden sanoille ei ole suomalaista vastinetta, joka tarkoittaisi täsmälleen sa-

9 Esim. Buckland 2017.

10 Lehdonvirta & Castronova 2014.

11 Flichy 2007.

12 Ks. esim. Floridi 2004; Bosancic & Matijevic 2019.

13 Haasio ym. 2019.

maa. Esimerkiksi tästä sopii Batesin¹⁴ tiivistelmä Beynon-Daviesin määritelmästä:

”Hänen mukaansa ”merkityksen toteutuminen” tapahtuu järjestelmissä kolmella ei tasolla, jotka ovat data, informaatio ja toiminta. Näihin viitataan myös käsitteillä forma, informa ja performa. Käsitteiden sisällöt viittaavat representaation, kommunikaation ja toteutuksen (*performance*) ideoihin.”¹⁵

Näin jo määritelmien kääntäminen muuttaa niiden sisältöä, ainakin sävyerojen verran.

Toinen haastava kysymys määrittelyssä on informaation ja *datan* välinen suhde. Floridille informaatio muodostuu datasta ja koko universumi on havainnoijista riippumatonta dataa, josta voi havaittaessa tulla informaatiota.¹⁶ Dataa on eroavaisuuksina ympäröivässä maailmankaikkeudessa (*dedomena*, ”*data in the wild*”), mutta dataa voivat myös olla erot signaalien tai symbolien välillä.¹⁷ Checklandille ja Holwellille tilanne on lähes vastaava, mutta datan ja informaation välissä on heille kuratoitu otos datasta (*capta*).¹⁸ Toisaalla data voi päinvastoin ja hämäävästi tarkoittaa joissakin yhteyksissä osuutta laajemmasta informaatiosta, osaa, jota sitten käytetään tutkimuksen pohjana. Informaatiotutkimuksessa tyypillisesti viitataan informaation arvoketjuun, jossa datasta saadaan informaatiota, joka jalostuu tiedoksi ja sitten tietämykseksi ja lopulta kenties viisaudeksi.¹⁹ Englanniksi sama tunnetaan myös nimellä DIKW (*data-information-knowledge-wisdom*).²⁰ Jalostus on monesti valikoivaa,

14 Bates 2018.

15 Beynon-Davies 2011. ”What he calls “the enactment of significance” consists in three levels of systems – data, information, and activity. These are called, respectively, forma, informa, and performa. Core to these three concepts are the ideas of representation, communication, and performance.”

16 Ks. esim. Floridi 2010; 2011.

17 Floridi 2010, 23–24.

18 Checkland & Holwell 1998.

19 Haasio & Savolainen 2004; ks. myös Niiniluoto 1996.

20 Bates 2018; Given ym. 2023.

minkä takia Galliers ja Newell sekä Ford toteavat, että datasta tulee informoivaa ihmisten omien linssien ja osaamisten kautta.²¹ Niin kutsutut tieto- tai informaatiojärjestelmät siis tosiasiaassa prosessoivat dataa, eivät informaatiota tai tietoa. Tiedon ja tietämyksen (kääntyvät yleensä kumpikin sanalla *knowledge*) ja viisauden (*wisdom*) voidaan katsoa olevan ihmisen mielessä, kun taas informaatio ja data ovat jotakin, joka on havaittavissa ja siirrettävissä. Informaatiosta tulee tietoa tulkinnan avulla, ja siksi tieto on subjektiivista. Malleissa, jotka tekevät eron tiedon ja tietämyksen välillä, tieto on informaatiota, jolla on konteksti, ja tietämys on tiedon tulkittamista sekä soveltamista.²² Viisaus taas on se, mitä tiedämme kyvyllämme käyttää tietämystämme, elämäkokemuksemme pohjalta.²³ Varsinkin monet varhaisista informaation määritelmistä perustuvat juuri tähän oletettuun eroon.²⁴

Usein sovellettu informaation yleismääritelmä (*general definition of information*, GDI) nojaa dataan ja merkitykseen. Siinä σ on (semanttista) informaatiota, jos ja vain jos se koostuu n -määrästä dataa niin, että n on yhtä suuri tai enemmän kuin yksi yksikkö, data on hyvin muodostettua (eli noudattaa käytössä olevan järjestelmän, koodin tai kielen sääntöjä) ja sillä on merkitys, joka voi olla, mutta jonka ei tarvitse olla, kielellinen.²⁵ Arkikielisemmin ja hiukan yksinkertaistaen voidaan todeta, että, jokin on GDI:n mukaan semanttista informaatiota, jos se koostuu riittävästä määrästä jonkin säännön mukaista dataa, jolla on merkitys.

Data voidaan jakaa esimerkiksi viiteen tyyppiin, joihin liittyy vastaavat tavat hahmottaa informaatiota.²⁶

21 Galliers & Newell 2003; Ford 2015.

22 Ks. Opetushallitus n.d.

23 Haasio ym. 2019.

24 Given ym. 2023.

25 Floridi 2011, 84.

26 Floridi 2010, 30–31.

- 1) Primääridata on ensisijaista ja yleensä tallennettua, esimerkiksi merkkejä tekstitiedostossa tai hälytysvalo vilkkumassa. Kun puhumme arkisemmin datasta tai informaatiosta, viittaamme yleensä tällaiseen.
- 2) Sekundäärinen data on sen vastine, joka tulee primääridatan puuttumisesta, eli esimerkiksi informoituminen siitä, että puhelin *ei* soi.
- 3) Metadata on dataa, joka kuvaa toista dataa ja metainformaatio informaatiota, joka kuvaa toista informaatiota. Ne voivat sisältää merkintöjä esimerkiksi datan sijainnista, muodosta tai päivityskäytännöistä.
- 4) Operationaalinen data ja operationaalinen informaatio käsittelevät informaatiojärjestelmien toimintaa.
- 5) Johdannaisdata ja -informaatio ovat sitä, mitä jokin toinen data jättää jälkeensä. Esimerkiksi bonuskortin kirjaamat ostotottumukset ilmaisevat myös, missä sen omistaja on luultavasti ollut milloinkin.

Lähtökohtaisesti meillä on kolme tapaa puhua informaatiosta: informaatio todellisuutena (esimerkiksi kuviot, puun renkaat ja sormenjäljet), informaatio todellisuutta varten (esimerkiksi algoritmit ja reseptit) ja informaatio todellisuutta koskien (esimerkiksi kuvaukset, kartat ja aikataulut).²⁷ Toisaalta voimme nähdä jaon informaatioon tietona (tietämistä välittävänä), informaation prosessina (informoiduksi tulemisena) ja informaation asiana (dokumentteina).²⁸ Monet tutkijat ovat tehneet kirjallisuuskatsauksia informaation käsitteestä, tuoreimpana Bates, joka listaa huolellisesti myös aiemmat katsaukset. Siksi tähän lukuun on valittu vain sellaisia määritelmiä, jotka joko ovat olleet erityisen vaikutusvaltaisia tai selvästi alleviivaavat käyttötapojen välisiä eroja.²⁹ Batesonin mukaan informaatio on “eroavaisuutta, joka tuottaa eroavuutta (tai: vaikut-

27 Floridi 2010, 74.

28 Buckland 2017, 24.

29 Bates 2018.

taa) jossakin myöhemmässä tilanteessa” (“difference which makes a difference in some later event”).³⁰ Brookes asettuu samoille linjoille ja näkee informaation asiana, joka tuottaa muutoksen vastaanottajansa tietorakenteissa.³¹ Dretskelle informaatio on (karkeasti ottaen) hyödyke, jonka on mahdollista tuottaa tietoa.³² Buckland puolestaan keskittyy enemmän informoitumiseen ja dokumentteihin kuin siihen, mitä itse informaatio on.³³ Monella tapaa onkin helpompaa viitata informaatioon informoiduksi tulemisen prosessina tai tapahtumana kuin määrittää varsinainen ydinkäsite.³⁴

Brier, pohjautuen Machlupin ajatuksiin, listaa joukon yleisiä oletuksia informaatiosta.³⁵ Oletukset ovat kiistanalaisia ja monella tapaa ongelmallisia, ja niihin on syytä suhtautua ennen kaikkea kuvauksina ajatuksista, joita informaation käsite voi ihmisissä herättää:

- Informaation tulisi koskea jotakin ennen tuntematonta.
- Informaation tulisi koskea jotakin, josta tiedetään vain vähän.
- Informaation tulisi vaikuttaa vastaanottajansa tietämyksen laajuuteen tai rakenteeseen.
- Informaation tulee koostua vain tulkitsemattomasta ”raa’asta” datasta.
- Informaation tulisi olla hyödyllistä.
- Informaatiota voidaan käyttää päätöksenteossa.
- Informaation tulisi vaikuttaa vastaanottajansa mahdollisiin toimiin.
- Informaation tulisi vähentää vastaanottajansa epävarmuutta.

30 Bateson 1972.

31 Brookes 1980.

32 Dretske 1981.

33 Buckland 1991; 2017.

34 Ks. esim. Bates 2018.

35 Machlup 1983; Brier 2008, 60.

- Informaation tulisi auttaa sanojen kontekstuaalisten merkitysten tunnistamisessa.
- Informaation tulisi muuttaa vastaanottajansa hyväksymiä asioita/oletuksia, erityisesti koskien suhtautumista (toiminta)mahdollisuuksiin.

Brier itsekin kyseenalaistaa nämä oletukset, sillä hän katsoo ne rajoittaviksi.³⁶ Ne eivät ole toimivia informaation määrittelemiseksi. On silti hyödyllistä tiedostaa ne, kun katsotaan eri määritelmiä sekä niiden historiaa ja konteksteja.

Informaatioteoreettinen lähestymistapa

Informaation kuvaamisen signaalin ymmärrettävänä tai merkityksellisenä osana voidaan katsoa alkaneen Nyquistin ja Hartleyn ajatuksista 1920-luvulla. Niin sanottu Hartleyn laki vuodelta 1928 määritteli informaation mitattavissa olevaksi asiaksi.³⁷ Tältä pohjalta informaatioteorian isänä pidetty Shannon loi teoriansa, joka mahdollisti sen tutkimisen, miten ja kuinka hyvin informaatio välittyy hälyisien kanavien läpi.³⁸ Siinä informaatio on merkkejä, jotka välitetään mahdollisimman tehokkaasti. Informaatioteorian kannalta niiden ei tarvitse olla merkityksellisiä tai muodostaa ymmärrettäviä kokonaisuuksia.

Jo tässä vaiheessa oli kuitenkin selvää, että teorianmuodostuksessa käytettiin sanaa informaatio toisin kuin arkikielessä. Samassa teoksessa Weaver toteaa:

”Sanaa informaatio käytetään tässä teoriassa erityisellä tavalla, jota ei tule sotkea sen tavallisen käyttöön. Ennen kaikkea, informaatiota ei tule sekoittaa merkitykseen. Kaksi viestiä, joista toinen on täynnä merkitystä ja toinen, joka sisältää pelkkää

³⁶ Brier 2008.

³⁷ Hartley 1928.

³⁸ Shannon & Weaver 1949.

älyttömyyttä, voivat tämän teorian perspektiivistä olla täysin samanarvoiset, mitä tulee informaatioon.”³⁹

Vaikka Shannonin teoria koskee ensisijaisesti teknisiä kommunikaatiojärjestelmiä, samaa ajatusta on mahdollista soveltaa myös muun muassa sosiaaliseen kommunikaatioon, jos sen pohjalta ajatellaan hälyn ja viestisisällön välistä suhdetta, esimerkiksi miettien, kuinka suuri osa kertojan puheesta välittyy kuulijalle hälyisässä ravintolassa. Tätä kautta informaatioteorian mekanistisillakin ajatuksilla on annettavaa informaatiotutkimuksen humanistisemmallekin laidalle, samalla kun se on varsin keskeinen sille osalle tieteenalaa, joka tutkii ihmisten ja tietojärjestelmien välistä vuorovaikutusta. Selkeyden vuoksi informaatioteoriassa puhutaan usein *Shannonin informaatiosta*, erotuksena muista, laajemmista informaation määritelmistä.

Informaatioteoreettisesta perspektiivistä informaatiolla on tyyppillisesti erilaisia vaiheita. Ne ovat *ilmaantuminen* (esimerkiksi löytäminen ja luominen), *välittäminen* (esimerkiksi verkottaminen, jakelu, käyttö, haku ja välitys), *käsittely ja hallinta* (esimerkiksi kerääminen, varmentaminen, muokkaus, organisointi, indeksointi, luokittelu, suodatus, päivitys, järjestely ja varastointi) ja *käyttö* (esimerkiksi seuranta, mallinnus, analysointi, selittäminen, suunnittelu, ennakointi, päätöksenteko, opastus, opetus ja oppiminen).⁴⁰ Tyyppillisenä kehänä askeleet ovat luominen/syntyy, kerääminen, tallennus, käsittely, välitys tai jakelu, käyttö ja lopulta kenties poistaminen, josta palataan alkuun.⁴¹ Vaikka kyseessä on vain esimerkki, informaatiolla on tyyppillisesti elinkaarensa, tarkoitimmepa käsitteellä mitä tahansa.⁴²

39 Shannon & Weaver 1949, 8–9 (suom. Harviainen.)

40 Floridi 2010, 4.

41 Floridi 2010, 5.

42 Haasio ym. 2019.

Informaation monet tasot

Badia jaottelee informaatiokäsitykset neljään luokkaan, joista kaikki viittaavat kaavoihin tai rakenteisiin, mutta eri tavoilla ja eri laajuudella.⁴³ Luokat ovat syntaktinen, semanttinen, pragmaattinen ja emergentti. Syntaktinen informaatio on koodin taso, jossa informaatiolla tarkoitetaan viestinnällisiä merkkejä. Sitä käyttivät esimerkiksi Shannon ja Kolmogorov. Merkkien tulkinta ja merkityksen luonti tapahtuu toisilla informaation tasoilla eli semanttisena tai pragmaattisena informaationa. Semanttisessa informaatiossa viestin merkit viittaavat johonkin eli syntyy merkitystä. Erilaisilla lisämääreillä voidaan rajata pois tulkintavaihtoehtoja, jolloin merkitys täsmentyy: ”sininen auto” kertoo enemmän kuin ”auto”, koska se rajaa muut värit pois).⁴⁴ Semanttinen informaatio voi olla joko opastavaa tai faktuaalista, eli se voi neuvoa, miten toimia tai kertoa asioiden tilasta – tai molempia.⁴⁵ Esimerkiksi huutava palohälytin sekä ohjeistaa vaihtamaan pariston että kertoo siitä, että paristo on lopussa, tai se kehottaa sammuttamaan tulipalon ja kertoo, että talo palaa. Semanttisen informaation teorioita on useita, tunnetuimpina niistä Bar-Hillelin ja Carnapin,⁴⁶ Dretsken⁴⁷ sekä Floridin⁴⁸ ajatukset. Siirryttäessä pois suoraan sanan informaatio käytöstä samoja kaikuja löytyy myös esimerkiksi hermeneutiikasta ja semiotiikasta, vain hiukan toisilla nimillä.

Pragmaattinen informaatio viittaa kontekstin huomioimiseen osana merkityksen rakentumista. Siinä, missä semanttinen taso keskittyy viittaussuhteisiin, pragmaattista informaatiokäsitystä edustavat teoriat pyrkivät ottamaan koko viitekehyksen huomioon. Tämä ero näkyy informaatiotutkimuksen alueella esimerkiksi eri tiedonhankinnan mallien tasoissa, joka ovat monimutkaisia. Keskiössä on

43 Badia 2019.

44 Haasio ym. 2019.

45 Floridi 2010.

46 Bar-Hillel & Carnap 1953.

47 Dretske 1981.

48 Esim. Floridi 2011.

se, mikä informaatio on tilanteessa oleellista. Pragmaattisen informaation haasteena tutkijalle on kuitenkin se, että muuttujia on mahdollista löytää lähes loputtomasti, joten koko kontekstin huomioiminen on käytännössä mahdotonta. Badian kategorioista viimeinen, emergentti informaatio, käsittelee informaatiota toimijoiden välisessä vuorovaikutuksessa, jossa syntyy yhteisvaikutuksena merkitystä, joka on suurempaa kuin vain osa-alueidensa semanttisen ja pragmaattisen tason summa.

Brenda Dervin käyttää jaottelussaan kolmea eri informaatiota: 1) informaatio viittaa ulkoiseen todellisuuteen, 2) informaatio viittaa sisäiseen todellisuuteen ja 3) informaatio sisältää kaikki ne tavat, joilla ihmiset merkityksellistävät maailmojaan.⁴⁹ Myös Floridi luettelee useita eri tasoja. Niistä viisi voidaan katsoa tarkemmaksi erittelyksi siitä, mitä Badia kutsuu syntaktiseksi informaatioksi.⁵⁰ Kuudentena Floridilla on viittauksellinen taso ja seitsemäntenä semanttinen. Floridille informaatio tarkoittaa kuitenkin käsitteenä ennen kaikkea semanttista informaatiota. Se on ”hyvin muodostettua, merkityksellistä ja totuudenmukaista dataa”.⁵¹ Oletus sisältää ajatuksen siitä, että se, mikä ei ole totuudenmukaista, ei ole informaatiota vaan pseudo-informaatiota. Tämä on ristiriidassa useiden muiden informaation määritelmien kanssa.⁵² Informaatio- ja viestintätutkimuksien puolella puhuisimme sen sijaan mis- ja disinformaatiosta, joista ensimmäinen on tahattomasti harhaanjohtavaa ja jälkimmäinen tarkoituksellisesti erehdyttävää, mutta molemmat kategoriat asettuvat silti informaation käsitteen sisään.⁵³ Myös Floridi hyödyntää näitä käsitteitä, vaikka sulkeekin ne varsinaisen *informaation* käsitteen ulkopuolelle.⁵⁴

49 Dervin 1977.

50 Floridi 2011; Badia 2019.

51 Floridi 2011, XIII.

52 Ks. esim. Bates 2018.

53 Karlova & Fisher 2013; Badia 2019; Savolainen 2022.

54 Floridi 2010.

Informaatio peliteoriassa

Kauppa- ja taloustieteiden peliteoriassa puhutaan täydellisestä ja epätäydellisestä informaatiosta.⁵⁵ Niistä ensimmäinen tarkoittaa tilannetta, jossa jollakin tai kaikilla osapuolilla on ainakin periaatteessa käytettävissään kaikki tilannetta koskeva informaatio päätöksentekoaan varten, mukaan lukien aiemmat tapahtumat. Epätäydellisessä puolestaan ei ole. Jos osalla on ja toisilla ei, puhutaan informaation epäsymmetriasta, jota käytetään myös kuvaamaan esimerkiksi sotatilanteita ja tietoturvallisuuden haasteita. Englanniksi tehdään vielä ero sen suhteen, onko informaatio kaiken kattavaa (*complete*) vai virheetöntä (*perfect*).

Täydellinenkin informaatio ei kuitenkaan takaa ratkaisujen virheettömyyttä, sillä esimerkiksi shakin tai gon potentiaalisten siirtojen määrä on niin suuri, ettei ihminen voi hahmottaa kaikkia mahdollisuuksia yhtä aikaa, vaikka hänellä periaatteessa on saatavillaan informaatio kaikesta paitsi vastustajan suunnitelmista.⁵⁶ Pokeri ja blackjack puolestaan edustavat hyvin erilaisia informaatio-olosuhteita, koska pokeriin kuuluu myös oleellisenä osana paitsi informaation epäsymmetria myös mis- ja disinformaatio.⁵⁷ Pelien ulkopuolisissa tilanteissa muuttujien määrä on vielä suurempi, joten informaatioepäsymmetria niissä kasvaa ja päädytään niin sanottuun rajattuun rationaalisuuteen sekä tyytymiseen (*satisficing*).⁵⁸ Tyytymisessä hyväksytään se, että saatavilla oleva informaatio ei ole parasta mahdollista, mutta se riittää päätöksenteon pohjaksi.

55 Esim. Rasmusen 2006; Von Neumann & Morgenstern 2007.

56 Pearce 1997.

57 Ks. esim. Manno 2020.

58 Simon 1981.

Viestinnällinen ja dokumenttipohjainen lähestymistapa

Shannon totesi aikanaan, että pelkästään jo informaatioteorian sisällä on lukuisia erilaisia käyttöjä sanalle informaatio.⁵⁹ Sen ulkopuolella määrä on vielä paljon suurempi ja ulottuu lukuisille tieteenaloille. Siksi on ollut tarpeen luoda informaatiotutkimuksen tarpeisiin työmääritelmiä, joissa sana määritellään vain käsillä olevan artikkelin tai väitöskirjan tarpeiden mukaan, yrittämättäkään luoda yleismaailmallista tai edes tieteenalan laajuista määrittystä. Esimerkiksi Harviainen ja Lieberoth⁶⁰ käyttävät Ingwersenin ja Järvelinin⁶¹ ajatusten pohjalta luotua työmääritelmää:

”Informaatio [...] on [tätä tutkimusta varten] määritelty potentiaalisesti viestisisällöksi, joka sisältyy mihin tahansa osaan dataa, ulottuen verbaalisista ilmaisuista fyysisiin esineisiin. Se valikoivasti joko hylätään tai liitetään osaksi tietorakenteita sellaisten henkilöiden toimesta, jotka hankkivat sitä tai kohtaavat sen.”

Edellä kuvattu määrittely tulee lähelle esimerkiksi Tague-Sutcliffen⁶² näkemystä, jossa henkilö voi informoitua dokumentista suhteessa siihen, millainen informaation tarve ja millaiset pohjatiedot hänellä on, sekä Bucklandin ajatuksia, joiden mukaan informaatio voi tarkoittaa informoitumisen prosessia, informaatiota prosessin tuloksena saatavana tietona tai informaationa asiana, joka voi mahdollistaa informoitumisen.⁶³ Myös esimerkiksi Thompsonin⁶⁴ ”organisoinnilla jalostettu tulos kokemuksen raaka-aineesta” tai Derrin⁶⁵ määritelmä ”abstraktista, merkityksellisestä representaatiosta

59 Shannon 1993, 180.

60 Harviainen & Lieberoth 2012.

61 Ingwersen & Järvelin 2005.

62 Tague-Sutcliffe 1995.

63 Buckland 2017.

64 Thompson 1968.

65 Derr 1985.

koskien esineestä [asiasta] tehtyjä päätelmiä” on tälle ajattelutavalle sukua, vaikka sen fokus onkin päätelmissä dokumentin informaatioisällön sijaan. Derrille informaatio on myös viestittävässä olevaa, informoivaa, toimintaa mahdollistavaa ja määrällistä siinä mielessä, että sen määrä voi vaihdella. Tällainen määrittelytapa on hyvin vastakkainen esimerkiksi Batesonin⁶⁶ ajatuksille tai Floridin⁶⁷ linjauksille, joissa jokseenkin kaikki on ainakin potentiaalisesti informaatiota ja jotka tulevat hyvin lähelle muun muassa alkuperäiskansojen maailmankuvia ja ajatusta ”oppimisesta maalta”, jolla on kaikki informaatio.⁶⁸ Toisaalta se laajentaa tiiviinkin määritelmän syntaktiselta tasolta ottamaan ainakin osan merkityksestä mukaan.

Huomattava joukko viestinnällisistä ja sosiaalisista informaation määritelmistä liittyy käytäntöteorioihin (ks. luku 2.1). Esimerkiksi Corneliuksen⁶⁹ mukaan informaatiota ei ole olemassa kuin sen tulokinnan mahdollistavissa käytännöissä ja Goguen⁷⁰ puolestaan toteaa, ettei informaatio voi olla koskaan täysin riippuvainen eikä täysin riippumaton kontekstista. Kontekstitekijöitä pohdittaessa on myös huomioitava, että informaatio voi olla joko normatiivista, jolloin se asettuu ympäröiviin normeihin, tai disnormatiivista, jolloin se poikkeaa niistä.⁷¹ Sama informaatio voi olla esimerkiksi alakulttuurin perspektiivistä normatiivista (esimerkiksi huumemyönteinen viestintä) mutta yhteiskunnan suunnasta katsottuna disnormatiivista. Disnormatiivisen informaation pohjalta muodostuu niin sanottua pimeää tietoa (*dark knowledge*), jolla tarkoitetaan sellaista tietämystä, joka on piiloteltua ja kiellettyä, koska sitä tai sen hankkimista pidetään moraalittomana, epäeettisenä tai laittomana.⁷² Tämä jaottelu on erillään informaation mahdollisesta totuusarvosta ja nähtävissä omana akselinaan tutkittaessa tietokäytäntöjä.

66 Bateson 1972.

67 Floridi 2011.

68 Ks. esim. Wilson 2008.

69 Cornelius 1996.

70 Goguen 1997.

71 Haasio 2015.

72 Burnett & Lloyd 2020.

Keskustelu

Kentän teoriat ja käsitteet eivät keskustelee kunnolla keskenään. Hyvänä esimerkkinä tästä on tapa, jolla Floridi toteaa, että informaatiotutkimus on sovellettua informaatiofilosofiaa, jättäen huomioimatta sen lukuisat muut kysymykset ja kontribuutiot ja rajaten informaation luonteen pohdinnan osaksi vain informaatiofilosofian oppialaa.⁷³ Taustalla on mitä ilmeisimmin oletus siitä, että informaatiotutkimus rajoittuisi vain kirjastojen tutkimukseen (tai kirjastotieteeseen) eikä ulottuisi laajemmalle. Näkökulmien välinen kuilu on silti ylitettävissä, jos esimerkiksi suomalaisesta perspektiivistä ajattelempa, että se, mitä meillä kutsutaan informaatiotutkimukseksi, sisältää sekä kirjastotieteen ja asiakirjahallinnan että myös osuuksia siitä, mitä Floridi kutsuu informaatiofilosofiaksi.

Vastaavasti informaatiotutkija Bates ohittaa Floridin näkemykset varsin kursorisesti luetellessaan erilaisia muiden (ja itsensä) laatimia määritelmiä.⁷⁴ Bawdenin ja Robinsonin mukaan Floridin informaatiofilosofia soveltuu mainiosti myös informaatiotutkimuksen perustaksi.⁷⁵ Floridi toisaalta toteaa, että perustavanlaatuisen kentän eli informaatiofilosofian on oltava valmis myös oppimaan luonteeltaan soveltavalta eli informaatiotutkimukselta kelvatakseen sen perustaksi.⁷⁶ Floridin mukaan näin myös tapahtuu. Kun katsomme informaatioteoriaan pohjautuvaa kirjallisuutta, Floridin oma tuotanto mukaan lukien, vuoropuhelu aidan siltä puolelta loistaa kuitenkin poissaolollaan.⁷⁷ Informaatiotutkimuksen informaatiokäsitys on siis jatkuvassa vaarassa joutua sulautetuksi mekanistisempiin malleihin, jotka suhtautuvat informaatioon pikemminkin mitattavana suureena kuin universaalina ilmiönä ja unohtavat Shannonin ja Weaverin huomiot termin monista merkityksistä. Tai ne saattavat

73 Floridi 2002; 2004.

74 Bates 2018.

75 Bawden & Robinson 2018.

76 Floridi 2002.

77 Esim. Floridi 2019.

Mackayn tavoin esittää, että myös sosiaalisempia merkityksiä voidaan mitata.⁷⁸

Johtopäätökset

Koska keskustelulla on ollut jo vuosikymmeniä aikaa sirpaloitua, on viisaampaa puhua yhteisistä juurista kuin yhteisestä käsitteestä. Alojen erilaisten tarpeiden takia on käytännössä pakko hyväksyä, että samalla sanalla on useita funktioita sekä useita rajapintoja muihin käsitteisiin kuten data ja tieto. Poikkirajaistakin yhdistelyä on mahdollista harjoittaa, kuten esimerkiksi Bawden ja Robinson ovat tehneet.⁷⁹ Samalla on kuitenkin tarpeen myös varoa, sillä kunkin version mukana tulee oma painolastinsa, joka ei välttämättä ole yhteensopiva toisen alan tavoitteiden ja menetelmien kanssa.

Hyvin harva informaatiotutkimuksen keskeisistä kirjoista vaihtuu määrittelemään informaation. Ne voivat kyllä puhua vaikkapa tavastaan määritellä informaation ja tiedon hankintaa tai hallintaa, mutta ne eivät kerro, mitä tarkoittavat sanalla informaatio. Sama tilanne on havaittavissa myös tietojärjestelmätieteen kirjallisuudessa. Voidaan siis hyvällä omallatunnolla väittää, että informaation määrittely on ulkoistettu toisille tieteenaloille, erityisesti informaatioteorian ja tietojenkäsittelyn haltuun. Ei näin ollen ole ihme, että tulokset ovat informaatiotutkimuksen tarpeisiin usein riittämättömät. Siksi informaatiotutkimukselle onkin erityisen tärkeää, että kirjoittajat käyttävät ainakin kriittisissä tapauksissa työmääritelmiä, joista näkee, mitä sanalla kyseissä kontekstissa tarkoitetaan. Näin sekä vältetään sekaannuksia että – jos hyvin käy – rikastutetaan koko tieteenalaa.

⁷⁸ Mackay 1969.

⁷⁹ Bawden & Robinson 2018.

Lähteet

- Badia, Anthony (2019) *The information manifold: Why computers cannot solve algorithmic bias and fake news*. Cambridge, MA: MIT Press.
- Bar-Hillel, Yehoshua & Rudolf Carnap (1953) Semantic information. *British journal for the philosophy of science* 4:14, 147–157.
- Bates, Marcia (2018) Information. Teoksessa John D. McDonald & Michael Levine-Clark (toim.) *Encyclopedia of library and information sciences*. Boca Raton: CRC Press, 2048–2063.
- Bateson, Gregory (1972) *Steps to an ecology of mind*. Chicago: University of Chicago Press.
- Bawden, David & Lyn Robinson (2018) Curating the infosphere: Luciano Floridi's philosophy of information as the foundation for library and information science. *Journal of documentation* 74:1, 2–17.
- Belkin, Nicholas J. (1978) Information concepts for information science. *Journal of documentation* 34, 55–58.
- Beynon-Davies, Paul (2011) *Significance: Exploring the nature of information, systems and technology*. Basingstoke: Palgrave Macmillan.
- Bosancic, Boris & Marta Matijevic (2019) Information as a construction. *Journal of librarianship and information science* 52:2, 620–630.
- Brier, Søren (2008) *Cybersemiotics: Why information is not enough!* Toronto: University of Toronto Press.
- Brookes, Bertram C. (1980) The foundations of information science: Part I: Philosophical aspects. *Journal of information science* 2:3–4, 125–133.
- Buckland, Michael (1991) *Information and information systems*. New York: Praeger.
- Buckland, Michael (2017) *Information and society*. Cambridge, MA: MIT Press.
- Burnett, Simon & Annemaree Lloyd (2020) Hidden and forbidden: Conceptualising dark knowledge. *Journal of Documentation* 76:6, 1341–1358.
- Case, Donald O. & Lisa Given (2016) *Looking for information: A survey of research on information seeking, needs, and behavior*. Bingley: Emerald.
- Checkland, Peter & Sue Holwell (1998) *Information, systems and information systems: Making sense of the field*. Chichester: Wiley.
- Cornelius, Ian (1996) Information and interpretation. Teoksessa Peter Ingwersen & Niels Ole Pors (toim.) *Integration in perspective, proceedings of CoLIS 2: Second international conference on conceptions of library and information science*; Royal School of Librarianship: Copenhagen, Denmark, October 13–16, 11–22.
- Derr, Richard L. (1985) The concept of information in ordinary discourse. *Information processing & management* 21:6, 489–499.
- Dervin, Brenda (1977) Useful theory for librarianship: Communication not information. *Drexel library quarterly* 13:3, 16–32.
- Dretske, Fred (1981) *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Flichy, Patrice (2007) *The internet imaginaire*. Cambridge, MA: MIT Press.
- Floridi, Luciano (2002) On defining library and information science as applied philosophy of information. *Social epistemology* 16:1, 37–49.

- Floridi, Luciano (2004) LIS as applied philosophy of information: A reappraisal. *Library trends* 52:3, 658–665.
- Floridi, Luciano (2010) *Information: A very short introduction*. Oxford: Oxford University Press.
- Floridi, Luciano (2011) *The philosophy of information*. Oxford: Oxford University Press.
- Floridi, Luciano (2019) *The logic of information*. Oxford: Oxford University Press.
- Ford, Nigel (2015) *Introduction to information behaviour*. Lontoo: Facet.
- Galliers, Robert D. & Sue Newell (2003) Back to the future: From knowledge management to data management. *Information systems and e-business management* 1:1, 5–13.
- Given, Lisa M. & Donald O. Case & Rebekah Willson (2023) *Looking for information: Examining research on how people engage with information*. Bingley: Emerald.
- Goguen, Joseph A. (1997) Towards a social, ethical theory of information. Teoksessa Geoffrey C. Bowker & Susan Leigh Star & William Turner & Les Gasser (toim.) *Social science research, technical systems, and cooperative work: Beyond the great divide*. Oxford: Oxford Psychology Press, 27–56.
- Haasio, Ari (2015) Disnormatiivinen ja normatiivinen informaatio. *Informaatiotutkimus* 34:4.
- Haasio, Ari & J. Tuomas Harviainen & Reijo Savolainen (2019) *Johdatus tiedonhankintatutkimukseen*. Helsinki: Avain.
- Haasio, Ari & Reijo Savolainen (2004) *Tiedonhankintatutkimuksen perusteet*. Helsinki: Avain.
- Hartley, Ralph V. L. (1928) Transmission of information. *Bell system technical journal* 7:3, 535–563.
- Harviainen, J. Tuomas & Andreas Lieberoth (2012) Similarity of social information processes in games and rituals: Magical interfaces. *Simulation & gaming* 43:4, 528–549.
- Ingwersen, Peter & Kalervo Järvelin (2005) *The turn: Integration of information seeking and retrieval in context*. Dordrecht: Springer.
- Karlova, Natascha A. & Karen E. Fisher (2013) A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Information research* 18:1.
- Lehdonvirta, Vili & Edward Castronova (2014) *Virtual economies: Design and analysis*. Cambridge, MA: MIT Press.
- Machlup, Fritz (1983) Semantic quirks in studies of information. Teoksessa Fritz Machlup & Una Mansfield (toim.) *The study of information: Interdisciplinary messages*. Chichester: Wiley, 641–671.
- MacKay, Donald M. (1969) *Information, mechanism, and meaning*. Cambridge, MA: MIT Press.
- Manno, Andrew (2020) *Toxic masculinity, casino capitalism, and America's favorite card game: The poker mindset*. Cham: Palgrave Macmillan.
- Niiniluoto, Ilkka (1996) *Informaatio, tieto ja yhteiskunta*. Helsinki: Edita.
- Opetushallitus (n.d.) Miten käsitteet, kuten 'data', 'informaatio' ja 'tieto' liittyvät toisiinsa? <https://www.oph.fi/fi/digiosaaminen/datatalousosaamisen-perusteita-perusopetukseen-ja-toiselle-asteelle/mihin-o> (haettu 19.12.2024).
- Pearce, Celia (1997) *The interactive book*. Indianapolis: Macmillan.

1.2 Informaation käsitteen eri merkitykset

- Rasmusen, Eric (2006) *Games and information: An introduction to game theory*. Chichester: Blackwell.
- Savolainen, Reijo (2022) Defending and refuting information sources rhetorically: The case of COVID-19 vaccination. *Journal of librarianship and information science* 55:4, 999–1014.
- Shannon, Claude E. (1993) *Collected papers*. Neil J. A. Sloane & Aaron D. Wyner (toim.) New York: IEEE Computer Society Press.
- Shannon, Claude E. & Warren Weaver (1949) *A mathematical theory of communications*. Urbana: University of Illinois Press.
- Simon, Herbert A. (1981) *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Tague-Sutcliffe, Jean (1995) *Measuring information*. San Diego: Academic Press.
- Thompson, Frederick B. (1968) The organization is the information. *American documentation* 19:3, 305–308.
- Von Neumann, John & Oskar Morgenstern (2007) *Theory of games and economic behavior: 60th anniversary commemorative edition*. Princeton: Princeton University Press.
- Wilson, Shawn (2008) *Research Is ceremony: Indigenous research methods*. Halifax: Fernwood.

Ihminen ja informaatio

2.1 Informaatiokäytännöt

J. Tuomas Harviainen

- Informaatiokäytännöt viittaavat jokseenkin vakiintuneisiin tapoihin, joilla ihmiset tunnistavat ja hyödyntävät erilaisia tiedonlähteitä ja kanavia sekä käyttävät ja jakavat niistä saamaansa informaatiota.
- Informaatiokäytäntöjä ovat esimerkiksi tiedon hankinta, käyttö, hallinta, jakaminen, torjunta ja luominen.
- Käytäntöteorioita ja niiden pohjalta muodostettuja informaatiokäytäntöteorioita on useita, mutta niitä yhdistää keskittyminen ihmisten pitkälti vakiintuneisiin toimintatapoihin erilaisissa konteksteissa.
- Informaatiokäytännöt toistuvat tyypillisesti suhteellisen samankaltaisina ja ovat siten tutkijoiden havaittavissa, mutta myös vaihtelevat toiminnan kontekstin mukaan ja voivat muuttua ajan tai olosuhteiden muutosten myötä.

Käytäntöteoriat ovat joukko teorioita, joilla voidaan ymmärtää ihmisten toimintaa eri konteksteissa.¹ Niiden pohjalta on muodostettu informaatiokäytäntöjen tutkimusta. Käsite informaatiokäytännöt (*information practices*) on hiljalleen tullut tutkimuksissa sanan tieto- tai informaatiokäyttäytyminen (*information behaviour*) rinnalle. Informaatio itse ei käyttäydy, vaan sitä hankitaan, muokataan, luodaan, käytetään ja jaetaan.² Nämä kaksi eri käsitettä mahdollistavat erilaisia tutkimuksia. Osalle tutkijoista on ollut selkeämpää keskittyä käytäntöihin eli toiminnan havaittaviin, toistuviin muotoihin sekä toiminnan sosiaaliseen rakentumiseen

¹ Schatzki 2001.

² Savolainen 2008.

kuin haastavammin määriteltävissä olevaan käyttäytymiseen.³ Informaatiokäyttäytymisellä on viitattu ihmisten käyttäytymiseen suhteessa informaatioon ja tietoon.⁴ Informaatiokäytännöt puolestaan viittaavat sellaisiin jokseenkin vakiintuneisiin tapoihin, joilla ihmiset tunnistavat ja hyödyntävät erilaisia tiedonlähteitä ja kanavia sekä käyttävät ja jakavat niistä saamaansa informaatiota.⁵ Niiden tutkimuksen keskiössä ovat toimijat, rutiinit ja prosessit tai, kuten strategisten käytäntöjen tutkija Richard Whittington toteaa, *practitioners, practices ja praxis*.⁶

Varsinkin 2000-luvun alkupuolen informaatiotutkimuksessa käyttäytymisen ja käytäntöjen käsitteitä käytetään myös limittäin.⁷ Lisäksi on tarjottu vaihtoehtona esimerkiksi *informaatiotoimintaa* (*information action, information activity*).⁸ Vedenjakajana selkeästi eriytettyihin linjauksiin voidaan pitää Reijo Savolaisen kirjaa *Everyday Life Information Practices* vuodelta 2008 ja sitä edeltänyttä artikkelia *Information behaviour and information practice*.⁹ Tiedonhaun ja -hakemisen (*information retrieval; information searching*) tutkimuksen puolella voidaan puhua myös informaatiovuorovaikutuksesta (*information interaction*), joka on osittain päällekkäinen sekä informaatiokäytäntöjen että -käyttäytymisen kanssa (aiheesta tarkemmin luvussa 2.3). Kaikkia näitä kattokäsitteitä ja niiden edustamia tutkimustapoja yhdistää se, että niissä yritetään hahmottaa, miten ihmiset toimivat suhteessa informaatioon.¹⁰ Siten tutkimustraditiot eivät ole toisensa poissulkevia ja niitä on edelleen mahdollista käyttää yhdessä, jos ne erityispiirteineen, vahvuuksineen ja rajoitteineen tuntuvat riittävän hyvin.¹¹ Merkityksellistä ei siis lopulta ole, onko yksi

3 Haasio ym. 2019.

4 Savolainen 2007a; 2007b; Wilson 1997; Given ym. 2023.

5 Savolainen 2008; Louhivuori 2019.

6 Whittington 2006; 2019.

7 Esim. Julien & Michaels 2001; McKenzie 2003; Talja & Hansen 2006.

8 Esim. Hartel 2006.

9 Savolainen 2007a; 2008.

10 Savolainen 2008.

11 Esim. Huuskonen 2014.

näistä käsitteistä toista parempi, toisen alaluokka vai jotakin rinnakkaista, vaan se, mitä kukin näistä linseistä antaa meidän ymmärtää.¹² Tässä luvussa keskitytään informaatiokäytäntöihin ja niiden erityispiirteisiin.

Informaatiokäytäntöjen määrittelyä hankaloittaa se, ettei ole olemassa yhtä käytäntöteoriaa vaan pikemminkin joukko samansuuntaisia teorioita ja metodologioita, jotka yhdessä muodostavat käytäntöteorioiden kentän.¹³ Niillä on kullakin omat painotuksensa, historiansa ja sanastonsa – sekä haasteensa. Niistä ei voi koota yhtä yhtenäistä teoriaa eikä edes yhteistä määritelmää sille, mitä käytännöt täsmälleen ovat, mutta ne ovat silti usein keskenään yhteensopivia.¹⁴ Informaatiotutkimuksen sovellutukset teorioista ovat ensisijaisesti keskittyneet kahden linjan soveltamiseen eli hyödyntämään varsinkin sosiologi Pierre Bourdieun (1930–2002) ja filosofi Theodore R. Schatzkin (1956–) ajatuksia. Näistä molempien taustalla on kuitenkin lukuisia muitakin teorioita erilaisissa vuorovaikutussuhteissa, minkä takia kentän hahmottamiseksi kattavasti on tarpeen lyhyesti katsoa myös muiden käytäntöteoreetikoiden näkemyksiä.

Käytäntöteoriat

Käytäntöteoriat nousevat ennen kaikkea kahdesta juuresta: mannermaisesta filosofiasta ja ranskalaisesta sosiologiasta.¹⁵ Varhaisia vaikutteita on otettu erityisesti Ludwig Wittgensteinilta ja Martin Heideggerilta, joiden ajatuksia jälkineen on havaittavissa lähes kaikilla myöhemmillä teoreetikoilla.¹⁶ Sosiologiassa erityisesti Pierre Bourdieu ja Anthony Giddens nostivat esiin asioita, jotka ovat sittemmin vakiintuneet käytäntöteorioiden ytimeen, erityisesti sen, miten enemmistö arkemme järjestymisestä pohjautuu käytäntöjen

12 Ks. esim. Savolainen & Thomson 2022.

13 Schatzki 2001.

14 Nicolini 2012.

15 Nicolini, 2012.

16 Wittgenstein 1953; Heidegger 1927/1962.

välisiin suhteisiin ja rakenteisiin.¹⁷ Taustahistoria on silti tosiasiaassa paljon näitä keskeisiä nimiä laajempi. Esimerkiksi sosiologian alakenttä etnometodologia, joka keskittyy tutkimaan niitä käytäntöjä, joiden kautta tietoa maailmasta muodostetaan, on tuonut käytäntöteorioihin useita tutkimusmenetelmiä.¹⁸

Teorioiden ja menetelmien hyödyntämisen yhteisvaikutuksena voimme puhua niin sanotusta käytäntöteoreettisesta käänteestä yhteiskuntatieteissä 2000-luvun alussa.¹⁹ Käytäntöteorioita sovelletaan toiminnan ymmärtämiseen myös monilla muilla aloilla, eli esimerkiksi informaatiokäytäntöjen tutkimus on ottanut ja yhä ottaa vaikutteita muiden tieteenalojen käytäntötutkimuksista, vaikka valitettavan harvoin pääsee itse vaikuttamaan niiden muiden kehitykseen.

Hyvä esimerkki rajapinnassa olevista käytäntöteorioista on Neuvostoliitossa alkunsa saanut toiminnan teoria, joka näkee ihmisen toiminnallisuuden muodostavan järjestelmiä, joiden osina ovat toimijat, toiminnan kohteet, toimintaa välittävät asiat, kuten työkalut ja merkit, sekä työn jakautuminen.²⁰ Aihetta on jatkokehitellyt Suomessa varsinkin Yrjö Engeström kollegoineen, tunnistuen myös säännöt ja yhteisöllisyyden osaksi toiminnan järjestelmiä.²¹ Käytäntöyhteisöjen tutkimuksen tavoin tämä kuvaa hyvin sitä, miten käytäntöteorioiden kirjo tarjoaa hyödyllisiä rajapintoja muihin tieteisiin, kuten tässä tapauksessa organisaatiotutkimukseen ja kasvatustieteisiin, mahdollistamalla ilmiöistä puhumisen samoilla käsitteillä, vaikka niitä ei aina olisikaan täsmälleen määriteltä.²²

Teorioiden ytimessä on se, että yksilöiden toiminta liittyy sosiaalisesti vakiintuneisiin käytäntöihin ja on niiden muokkaamaa. Käytännöt ovat ikään kuin alakulttuureja, jotka vaikuttavat tapahtumien pintojen alla.²³ Ne voidaan määritellä esimerkiksi ”joukoksi

17 Bourdieu 1977; 1990; Giddens 1984.

18 Esim. Garfinkel 1967; Lynch 2001.

19 Reckwitz 2002; Schatzki ym. 2001.

20 Leont’ev 1978.

21 Engeström 1995; Engeström ym. 1999.

22 Ks. esim. Lave & Wenger 1991; Wenger 1998.

23 Vaara & Whittington 2012.

organisointuneita tekemisiä tai sanomisia, tehtäviä ja projekteja” tai ”hyväksytyiksi tavoiksi tehdä asioita, kehollisina ja materiaalien välittämistä, jotka jaetaan eri toimijoiden kesken ja jotka rutinoituvat ajan myötä”.²⁴ Ne myös ruokkivat omaa jatkuvuuttaan, vaikka ne voivat muuttua ajan kuluessa ja uusia käytäntöjä voi syntyä entisten kadotessa.²⁵ Étienne Wengerin mukaan käytännöt ovat aina sekä historiallisessa että sosiaalisessa viitekehyksessä tekemistä, mikä antaa rakenteen ja merkityksen toiminnallemme.²⁶ Oleellista käytännöille on siis vakiintuneisuus ja sosiaalisuus eli jaettuus ryhmissä tai kulttuureissa. Siksi käytäntöteoreettinen lähestymistapa alleviivaa kontekstin merkitystä, mikä näkyy myös tutkittaessa informaatiokäytäntöjä.

Nicolinin mukaan viisi asiaa erottavat käytäntöteoriat muista lähestymistavoista:²⁷

- 1) Toiminnan merkitys painottuu keskeisenä kaikkien sosiaalisen elämän aspektien, myös sosiaalisten rakenteiden, luomisessa ja jatkuvuudessa.
- 2) Kehoilta ja muilla materiaalisilla asioilla on kriittisen tärkeä rooli kaikessa sosiaalisessa, vaikka monet käytännöistä luodaan ja ylläpidetään erityisesti erilaisten vuoropuheluiden kautta.
- 3) Yksilöillä on omat roolinsa osana käytäntöjen verkostoja, koska jokainen konteksti vaatii soveltamista.
- 4) Käytäntökeskeisyys pakottaa toisenlaiseen suhteeseen tiedon ja merkitysten kanssa. Käytäntöteorioiden perspektiivistä tieto on kykyä toteuttaa sosiaalisia ja materiaalisia toimintoja – opittua, kehollista, kenties ohjeisiin tallennettua. Osallisuus olemassa olevista

24 Schatzki, 2002, 73; Vaara & Whittington 2012.

25 Knorr Cetina, 2001.

26 Wenger 1998.

27 Nicolini 2012, 3–8.

käytännöistä vaatii monenlaista oppimista sekä merkitysten ja usein myös arvojen sisäistämistä.

- s) Käytännöt ovat sidoksissa valtaan, politiikkaan ja konflikteihin. Mahdollistaessaan toimintaa ne myös sulkevat joiltakin ihmisiltä tai asioilta mahdollisuuksia toimia.

Käytäntöteoriat painottavat toimijoiden välisiä suhteita ja siten hylkäävät ajatuksen siitä, että ilmiöt olisivat esimerkiksi vain mikro- tai makrotasolla, joko yksilökohtaisia tai pelkästään yhteisöllisiä. Huomio käytäntöjen materiaalisuudesta puolestaan erottaa nämä tutkimustraditiot ajatuksista, joiden näkemyksen mukaan todellisuutta luodaan kokonaan vuoropuheluilla tai symbolisella vuorovaikutuksella. Informaatiotutkimuksen kannalta erityisen hedelmällisiä tarttumapintoja on varsinkin niissä teorioista, jotka keskittyvät vuoropuheluihin ja/tai kieleen tai materiaaleihin tallennettuun informaatioon osana laajempaa keskustelua käytäntöjen materiaalisuudesta ja ei-inhimillisten toimijoiden rooleista niissä. Niitä on kuitenkin hyödynnetty vielä varsin vähän informaatiota tutkittaessa.²⁸

Informaatiokäytännöt

Informaatiokäytännöt ovat yksi alalaji (tai oikeastaan joukko sellaisia) inhimillisten käytäntöjen piirissä. Niillä on selkeitä erityispiirteitä. Informaatiolla on lisäksi usein keskeinen rooli myös niissä käytännöissä, joita tutkittaessa ei puhuta informaatiosta, koska tietämys on silti niissä oleellisessa osassa (ks. luku 1.2).²⁹ Ian Cornelius menee asiassa vielä yhden pykälän pitemmälle ja väittää, ettei informaatiota erillisenä asiana edes ole olemassa käytäntöjen ulkopuolella.³⁰

Reijo Savolaiselle informaatiokäytännöt ovat “joukko sosiaalisesti ja kulttuurisesti vakiintuneita tapoja tunnistaa, hankkia, käyt-

28 Esim. Shatzki 2017; Shove 2017; Louhivuori 2019.

29 Cox 2012.

30 Cornelius 1996.

tää ja jakaa informaatiota”³¹ Taustalla on ennen kaikkea Bourdieun ajatus *habituksista*, joka syntyy kehollisesti omaksutuista preferensseistä, havainnointitavoista ja tavoista reagoida asioihin.³² Yksilö sisäistää ne sosiaalisista ja kulttuurisista tavoista ja ajattelujärjestelmistä. Kohdatessaan eri olosuhteita (”kenttiä”) habitus tuottaa henkilön tilanteen eli hänelle kertyneen pääoman mahdollistamia käytäntöjä. Chun Wei Choo puhuu informaatiokäytännöistä, jotka ovat toistuvia käyttäytymismalleja, jotka vahvistavat rooleja, rakenteita ja vuorovaikutuksen muotoja organisaatioissa ja jotka ovat havaittavissa ihmisten tavoissa toimia suhteessa informaatioon.³³

Pamela J. McKenzie toteaa, että siirtyminen käyttäytymisestä käytäntöihin vaihtaa tutkimuksen painopisteen kognitiivisesta sosiaalisesta ja mahdollistaa tiedonhankkijoiden tutkimisen heidän sosiaalisissa konteksteissaan.³⁴ Se huomioi silti samat asiat kuin tietokäyttäjien tutkimus ja antaa lisäksi tilaisuuden tutkia tilanteita, jossa informaatiota kohdataan tai vastaanotetaan jonkun muun kuin sen hankkijan aloitteesta.³⁵ Tietokäytännöt voivat olla niiden toteuttajan osalta aktiivisia tai passiivisia. Esimerkiksi McKenzien klassisen mallin mukaan tyypillisiä informaatiokäytäntöjen muotoja ovat aktiivinen hankinta (*active seeking*), aktiivinen läpikäynti (*active scanning*), epäsuora seuranta (*non-directed monitoring*) ja hankinta toisten avulla (*information seeking by proxy*).³⁶

Jaotteluja on kuitenkin mahdollista tehdä lukuisia, lähtien tiedontarpeiden tiedostamisesta ja päätyen informaation ja tiedon jakamiseen, käyttöön, pimittämiseen, torjuntaan ja jopa hävittämiseen.³⁷ Informaatiotutkimus on perinteisesti ollut kiinnostunut näistä ennen kaikkea tiedonhankinnasta, mikä näkyy siihen liitty-

31 Savolainen 2008, 2.

32 Bourdieu 1984; 1990.

33 Choo 2007.

34 McKenzie 2003.

35 Ks. esim. Erdelez 1999.

36 McKenzie 2003.

37 Gorichanaz 2019; Huvila 2022.

vien tutkimusten määrässä.³⁸ Toinen kiinnostuksen kohde on ollut tiedon hallinta, sekä henkilökohtaisella että organisaatioiden tasolla. Vastaavasti kauppatieteissä on vahva traditionsa tiedon jakamisen (*knowledge sharing, knowledge transfer*) tutkimukselle ja design-tutkimuksessa informaation kulun haasteiden ymmärtämiselle.³⁹ Ne sopivat yleensä hyvin yhteen informaatiotutkimuksen havaintojen kanssa ja ovat siten helposti tieteenalamme hyödynnettävissä.

Informaatiokäytäntöjen tutkimuksessa on kuitenkin kasvavissa määrin tehty uusia aluevaltauksia ja kiinnostuttu kentän tutkimustarpeista. Viime vuosina varsinkin informaation jakamisen tutkiminen on ottanut harppauksia eteenpäin, pitkälti siksi, että internetin ja sosiaalisten medioiden yleistyminen on tuonut monia sen ilmiöistä aiempaa näkyvämmiksi. Vaikka informaatiokäytäntöjen vakiintuneisuus korostaa herkästi toiminnan rationaalisia piirteitä, niillä on usein myös vahvat affektiiviset, emotionaaliset puolensa, jotka vaikuttavat tehtäviin ratkaisuihin.⁴⁰ Ratkaisuihin vaikuttavat myös erilaiset ajattelun oikotiet sekä päätöksenteon vinoumat.⁴¹

Monet keskeiset informaatiokäytäntöjen tutkimukset huomioivat näitä aspekteja. Tällainen on esimerkiksi Annemaree Lloydin ja Michael Olssonin artikkeli autoja restauroivien ihmisten tietokäytännöistä.⁴² Se sisältää myös kattavan tiivistelmän käytäntöteoreettiseen lähestymistapaan. Lloyd ja Olsson puhuvat käytäntöjen lisäksi informaatioympäristöistä (*information environments*), jotka määrittävät tavan toteuttaa käytäntöjä siten, että ne liittävät saman aiheen parissa olevia toimijoita yhteisiin viitekehyksiin ja tarjoavat vakaata tietoa hyödynnettäväksi. Informaatioympäristöissä yhdistyy fyysinen, sosiaalinen ja kognitiivinen informaatio.⁴³ Informaatiomaisemat (*information landscapes*) puolestaan kertovat informaatioym-

38 Haasio & Savolainen 2004; Haasio ym. 2019; Given ym. 2023.

39 Esim. Choo 2006; Nonaka & Takeuchi 1995; Eckert ym. 2001; Deken ym. 2012.

40 Esim. Olsson 2013.

41 Ks. Honkanen 2016; Thaler 2015.

42 Lloyd & Olsson 2019.

43 Lloyd 2007.

päristöjen pohjalta, miten asiat tapahtuvat, ja kytkeytyvät tiettyjen tilanteiden ihmisiin, esineisiin ja kehollisiin toiminnallisuuksiin.⁴⁴ Yhteisöjen uudet jäsenet oppivat niiden käytännöt kartoittamalla tiedonlähteitä informaatioympäristössä. Oppiessaan he rakentavat informaatiomaisemansa toimintatavat sekä jaetulla informaatiolla että tarkkaillen muita.

Nicolinin mukaan käytäntöteoreettisessa lähestymistavassa ensisijaiseksi asetetaan käytäntö, ei toimijoita, sillä vasta, kun hahmotamme käytännöt jossakin tilanteessa, voimme alkaa katsoa, millä tavoin ne vaikuttavat toimijoiden toimintaan ja mahdollisesti rajoittavat tai muokkaavat sitä.⁴⁵ Tämä pätee myös suureen osaan informaatiokäytäntöjen tutkimuksesta. Toisaalta, kuten Savolainen toteaa, kyseessä on yksilön toiminta, joka pohjautuu hänen elämäntapaansa (*Way of Life*), joka puolestaan on käytäntöjen muokkaama.⁴⁶ Tällaisten ajatusten välinen tasapaino on löydettävissä Annemaree Lloydilta, joka kuvatessaan informaatiolukutaitoja Schatzkin käytäntöteorian pohjalta samalla avaa niiden tausta-ajattelua ja osoittaa, miten käytäntölähtöinen tutkimus voi mahdollistaa uusia perspektiivejä informaation ilmiöihin.⁴⁷ Parhaimmillaan käytäntöjen tutkimuksella päästään lähelle aihetta ja ymmärretään sekä kontekstin että yksilön ja sosiaalisten aspektien osuudet tilanteessa. Tämä on tärkeää, sillä kuten jo Bourdieu totesi, habituksen ja olosuhteiden yhteisvaikutus tuottaa sen, etteivät käytännöt ole koskaan täysin identtisiä tilanteesta toiseen, vaikka ne ovatkin pohjimmiltaan vakiintuneita.⁴⁸ Muutamme toimintaamme kunkin kontekstin mukaan, myös tiedon hankinnassa, jakamisessa ja käytössä, mutta teemme niin yleensä silti vakiintuneiden toimintatapojemme pohjalta.⁴⁹

44 Lloyd 2010.

45 Nicolini 2012.

46 Savolainen 1995.

47 Lloyd 2010.

48 Bourdieu 1990.

49 Schatzki 2002.

Informaatiokäytäntöjen tutkimuksesta

Informaatiokäytäntöjen tutkimus on, kuten käytäntöteoriat laajemminkin, monien lähestymistapojen joukko, ei yhtenäinen traditio.⁵⁰ Informaatiokäytännöissä on huomattavan pitkälti painotuttu tunnetuimpiin käytäntöteorioihin. Informaatiokäytäntöjen tutkimuksessa on myös monesti keskitytty mikrotason ilmiöihin (esimerkiksi yksilöön jossakin tilanteessa) tai paikalliseen kontekstiin, vaikka teoriakehukset eivät sitä velvoitakaan.

Teorioiden mukaan käytännöt tyypillisesti muodostavat laajempia kokonaisuuksia, joissa eri käytännöt yhdistyvät verkostoiksi ja vaikuttavat toisiinsa sekä ovat riippuvaisia materiaalisista järjestelyistä ja esineistä.⁵¹ Siksi on tärkeää pitää mielessä, että käytäntöjen jakaminen usein ylittää yksittäisen organisaation tai alakulttuurin rajat. Käänteisesti tästä myös seuraa, että ymmärtääksemme yhden ryhmän tai jopa yksilön käytäntöjä meidän on katsottava asiaa paljon laajemmin kuin vain hänen välittömän kontekstinsa kannalta. Taustalla voi olla useita kulttuurisia konventioita. Toisaalta nämä laajat yhteisesti jaetut käytännöt kokevat jatkuvasti mahdollisia muutoksia ja säättöjä paikallisella tasolla, kun niitä sovelletaan ja sovitaan tilannekohtaisesti konteksteihin.⁵² Parhaimmillaan informaatiotutkimus ottaa nämä kaikki huomioon ja kykenee informaatiokäytäntöihin keskittyessään ongelmitta osoittamaan, miten eri tasot ja sosiaalinen viitekehys mahdollistavat yksilön toimintaa. Esimerkkejä tästä ovat Jenna Hartelin tutkimus gourmet-kokeista sekä jo aiemmin mainittu Lloydin ja Olssonin artikkeli autoharrastajista.⁵³

Näissä korostuu jälleen käytäntöjen materiaallinen puoli, joka on usein monen tekijän summana jotakin, joka rakentuu ympäröivän yhteisön tai yhteiskunnan tarjoamista mahdollisuuksista enemmän

50 Pilerot ym. 2017.

51 Yhdistelmille on eri teorioissa useita eri nimiä, ja niiden sisällöissä on sävyeroja. Ne ovat esimerkiksi ”kenttiä” (Bourdieu 1990), ”sosiaalisia järjestelmiä” (Giddens 1984) tai ”paikkoja” (Schatzki 2002).

52 Vaara & Whittington 2012.

53 Hartel 2006; Lloyd & Olsson 2019.

kuin toiminnan (tai tiedontarpeen) omista lähtökohdista.⁵⁴ Informaatiotutkimuksen perspektiivistä olisi mahdollista siis tutkia uusilla tavoilla esimerkiksi ihmisten ja dokumenttien välistä vuorovaikutusta, joka on jo pitkään ollut yksi oppiaineen ydinaiheista.

Informaatiokäytäntöjä on tutkittu myös organisaatioissa ja keskustelua on sittemmin jatkettu informaatiotutkimuksen puolelle. Esimerkiksi Marchandin ja kollegoiden jaottelua informaatiokäytännöistä esiintymässä tiedonhallinnan, tietojärjestelmien tai informaatiokulttuurin alueilla on laajennettu katsomaan myös tiedon jakamista.⁵⁵

Käytäntöteorioiden hyödyistä

Informaatiokäytäntöjen suhteellinen vakaus tekee niistä hyvän tutkimuskohteen, vaikka käytännöt myös muuttuvat. On mahdollista tunnistaa projekteja, joissa informaatiota hankitaan, luodaan, jaetaan ja kenties myös pidätellään. Keskittyminen käytäntöihin osoittaa myös askelia, joilla käytännöt kytkeytyvät toisiinsa. Esimerkiksi informaation luomista tehdään usein sen jakamista varten ja se vaatii tiedonhallintaa mahdollista jatkokäyttöään varten.⁵⁶

Käytäntöteoreettinen lähestymistapa velvoittaa tutkijansa kiinnittämään huomiota konteksteihin ja sosiaalisiin verkostoihin, koska käytännöt ovat riippuvaisia toisista käytännöistä, konventioista, kulttuurisista oletuksista, kehoista ja materiasta. Ne tapahtuvat vakiintuneina osina historiallisia jatkumoina mutta muokautuvat kunkin kontekstin mukaan. Mikään ei luonnollisestikaan estä tutkimasta samoja asioita osana informaatiokäyttämistä tai -vuorovaikutusta, mutta käytäntöteoriat antavat tutkimukselle sanastoa ja kieliopin, joilla jäsentää havaittua, sekä menetelmiä, joilla kiinnittää erityistä huomiota juuri informaatiokäytäntöjen syntyyn, toistuvuuteen ja siihen, miten niitä jaetaan ja ylläpidetään.

54 Shove 2017.

55 Marchand ym. 2000; Zárraga-Rodríguez & Álvarez 2015.

56 Esim. Harviainen & Melkko 2022; McKenzie ym. 2014.

Lähteet

- Bourdieu, Pierre (1977) *Outline of a theory of practice*. Cambridge: Cambridge University Press.
- Bourdieu, Pierre (1984) *Distinction: A social critique of the judgement of taste*. Cambridge, MA: Harvard University Press.
- Bourdieu, Pierre (1990) *The logic of practice*. Cambridge: Polity Press.
- Choo, Chun Wei (2006) *The knowing organization. How organizations use information to construct meaning, create knowledge and make decisions*. New York: Oxford University Press.
- Choo, Chun Wei (2007) Information seeking in organizations. Epistemic contexts and contests. *Information research* 12:2, paper 298.
- Cornelius, Ian (1996) Information and interpretation. Teoksessa Peter Ingwersen & Niels Ole Pors (toim.) *Integration in perspective, proceedings of CoLIS 2: Second international conference on conceptions of library and information science*; Royal School of Librarianship: Copenhagen, Denmark, October 13–16, 11–22.
- Cox, Andrew M. (2012) An exploration of the practice approach and its place in information science. *Journal of information science* 38:2, 176–188.
- Deken, Fleur & Maaike Kleinsmann & Marko Aurisicchio & Kristina Lauche & Rob Bracewell (2012) Tapping into past design experiences: Knowledge sharing and creation during novice–expert design consultations. *Research in engineering design* 23, 203–218.
- Eckert, Claudia & P. John Clarkson & Martin Stacey (2001) Information flow in engineering companies: Problems and their causes. Teoksessa *Proceedings of the 13th international conference on engineering design: Design management – Process and information issues*, 43–50.
- Engeström, Yrjö (1995) Kehittävä työntutkimus: perusteita, tuloksia ja haasteita. Helsinki: Hallinnon kehittämiskeskus.
- Engeström, Yrjö & Reijo Miettinen & Raija-Leena Punamäki (1999) *Perspectives on activity theory*. Cambridge: Cambridge University Press.
- Erdelez, Sanda (1999) Information encountering: It's more than just bumping into information. *Bulletin of the American Society for Information Science* 25:3, 25–29.
- Garfinkel, Harold (1967) *Studies in ethnomethodology*. Englewood Cliffs: Prentice Hall.
- Giddens, Anthony (1984) *The constitution of society*. Cambridge: Polity Press.
- Given, Lisa M., Donald O. Case & Rebekah Willson (2023) *Looking for information: Examining research on how people engage with information*. Bingley: Emerald.
- Gorichanaz, Tim (2019) Information creation and models of information behavior: Grounding synthesis and further research. *Journal of librarianship and information science* 51:4, 998–1006.
- Haasio, Ari & J. Tuomas Harviainen & Reijo Savolainen (2019) *Johdatus tiedonhankintatutkimukseen*. Helsinki: Avain.
- Haasio, Ari & Reijo Savolainen (2004) *Tiedonhankintatutkimuksen perusteet*. Helsinki: Avain.

- Hartel, Jenna (2006) Information activities and resources in an episode of gourmet cooking. *Information research* 12:1, paper 282.
- Harviainen, J. Tuomas & Runo Melkko (2022) Organizational information creation through a design game: A sensemaking perspective. *Library & information science research*, 44:3, 101172.
- Heidegger, Martin (1962) *Being and time*. Oxford: Blackwell.
- Honkanen, Henry (2016) *Vaikuttamisen psykologia: Mielen muuttamisen tiede ja taito*. Helsinki: Arena-Innovation Oy.
- Huuskonen, Saila (2014) *Recording and use of information in a client information system in child protection work*. Tampere: University of Tampere.
- Huvila, Isto (2022) Making and taking information. *Journal of the association for information science and technology* 73:4, 528–541.
- Julien, Heidi & David Michels (2001) Source selection among information seekers: Ideals and realities. *Canadian journal of information and library science* 25:1, 1–18.
- Knorr Cetina, Karin (2001) Objectual practice. Teoksessa Theodore R. Schatzki & Karin Knorr-Cetina & Eike von Savigny (toim.) *The practice turn in contemporary theory*. New York: Routledge, 175–188.
- Lave, Jean & Étienne Wenger (1991) *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Leont'ev, Aleksei N. (1978) *Activity, consciousness, personality*. Englewood Cliffs: Prentice Hall.
- Lloyd, Annemaree (2007) Recasting information literacy as sociocultural practice: Implications for library and information science researchers. *Information research* 12.
- Lloyd, Annemaree (2010) Framing information literacy as information practice: Site ontology and practice theory. *Journal of documentation* 66:2, 245–258.
- Lloyd, Annemaree & Michael Olsson (2019) Untangling the knot: The information practices of enthusiast car restorers. *Journal of the association for information science and technology* 70:12, 1311–1323.
- Louhivuori, Jaakko (2019) Informaatiokäytäntöjen tutkimuksen teoreettisia ja metodologisia mahdollisuuksia: toimijaverkkoteoria ja käytäntölähtöinen tutkimusote. *Informaatiotutkimus* 38:3–4, 7–19.
- Lynch, Michael (2001) Ethnomethodology and the logic of practice. Teoksessa Theodore R. Schatzki & Karin Knorr-Cetina & Eike von Savigny (toim.) *The practice turn in contemporary theory*. New York: Routledge, 131–148.
- Marchand, Donald A. & William J. Kettinger & John D. Rollins (2000) Information orientation: People, technology and the bottom line. *Sloan management review* 41:4, 69–80.
- McKenzie, Pamela J. (2003) A model of information practices in accounts of everyday-life information seeking. *Journal of documentation* 59:1, 19–40.
- McKenzie, Pamela J. & Elisabeth Davies & Sherilyn Williams (2014) Information creation and the ideological code of managerialism in the work of keeping track. *Information research* 19:2, paper 614.
- Nicolini, Davide (2012) *Practice theory, work & organization: An introduction*. Oxford: Oxford University Press.

- Nonaka, Ikujiro & Hirotaka Takeuchi (1995) *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.
- Olsson, Michael (2013) Gently to hear, kindly to judge: The affective information practices of theatre professionals and journalists. *Information research* 18:3, paper C22.
- Pilerot, Ola & Björn Hammarfelt & Camilla Moring (2017) The many faces of practice theory in library and information studies. *Information research* 22:1, CoLIS paper 1602.
- Reckwitz, Andreas (2002) Toward a theory of social practices: A development in cultural theorizing. *European journal of social theory* 5:2, 243–263.
- Savolainen, Reijo (1995) Everyday life information seeking: Approaching information seeking in the context of way of life. *Library & information science research* 17:3, 259–294.
- Savolainen, Reijo (2007a) Information behaviour and information practice; Reviewing the “umbrella concepts” of information seeking studies. *Library quarterly* 77:2, 109–132.
- Savolainen, Reijo (2007b) Motives for giving information in non work contexts and the expectations of reciprocity. The case of environmental activists. *ASIST 2007. Proceedings of the annual meeting of the American society for information science and technology, Milwaukee, Wisconsin, October 19–24, 2007*, 1–13.
- Savolainen, Reijo (2008) *Everyday information practices: A social phenomenological perspective*. Lanham: Scarecrow Press.
- Savolainen, Reijo & Leslie Thomson (2022) Assessing the theoretical potential of an expanded model for everyday information practices. *Journal of the association for information science and technology* 73:4, 511–527.
- Schatzki, Theodore (2001) Practice mind-ed orders. Teoksessa Theodore R. Schatzki & Karin Knorr-Cetina & Eike von Savigny (toim.) *The practice turn in contemporary theory*. New York: Routledge, 42–55.
- Schatzki, Theodore R. (2002) *The site of the social*. University Park: Pennsylvania State University Press.
- Schatzki, Theodore (2017) Sayings, text and discursive formations. Teoksessa Allison Hui & Theodore R. Schatzki & Elizabeth Shove (toim.) *The nexus of practices: Connections, constellations, practitioners*. Oxford: Routledge, 126–140.
- Schatzki, Theodore R. & Karin Knorr-Cetina & Eike von Savigny (toim.) (2001) *The practice turn in contemporary theory*. New York: Routledge
- Shove, Elizabeth (2017) Matters of Practice. Teoksessa Allison Hui & Theodore R. Schatzki & Elizabeth Shove (toim.) *The nexus of practices: Connections, constellations, practitioners*. Oxford: Routledge, 155–168.
- Talja, Sanna & Preben Hansen (2006) Information sharing. Teoksessa Amanda Spink & Charles Cole (toim.) *New directions in human information behavior*. Dordrecht: Springer, 113–134.
- Thaler, Richard H. (2015) *Väärin käyttäytyminen: Käyttäytymistaloustieteen synty*. Helsinki: Terra Cognita.
- Vaara, Eero & Richard Whittington (2012) Strategy-as-practice: Taking social practices seriously. *Academy of management annals* 6:1, 1–52.

- Wenger, Étienne (1998) *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.
- Whittington, Richard (2006) Completing the practice turn in strategy research. *Organization studies* 27:5, 613–634.
- Whittington, Richard (2019) *Opening strategy. Professional strategists and practice change, 1960 to today*. Oxford: Oxford University Press.
- Wilson, Thomas D. (1997) Information behavior: An inter-disciplinary perspective. *Information processing & management* 33:4, 551–572.
- Wittgenstein, Ludwig (1953) *Philosophical investigations*. Oxford: Basil Blackwell.
- Zárraga-Rodríguez, Marta & Maria Jesús Álvarez (2015) Information practices and competences that evidence information capability in a company. *Information research* 20:2, paper 664.

2.2 Lukutaitojen ja lukemisen tutkimus

Sanna Talja

- Tarkastelun kohteena ovat lukemisen ja lukutaidon käsitteet monilukutaitoteorian näkökulmasta.
- Monilukutaitoajattelua esitellään sekä sosiokulttuuriseen teoriaperinteeseen pohjautuvana teoriana että pedagogisena käytäntönä, joka korostaa laajaa tekstikäsitystä.
- Sosiokulttuuriseen teoriaperinteeseen nojaava lukemisen tutkimus suosii etnografista tutkimusotetta kokeellisen ja lukutaitoja testaavan tutkimusotteen sijaan.
- Kiinnostuksen keskiössä ovat todelliset lukijat käytäntöineen ja kokemuksineen, eivät niinkään kirjalliset tekstit tai tiedonlähteet ja tiedonhankintakanavat.

Lukutaitojen tukeminen ja kehittäminen on yhteiskunnallisessa keskustelussa asia, joka puhuttaa ja huolestuttaa jatkuvasti. Digitalisaation myötä kansalaisilta vaaditaan aina vain enemmän kykyjä ja valmiuksia toimia erilaisissa teknologisissa ja mediaympäristöissä toteuttaakseen arkipäiväisiä perusasioitaan. Tieto- ja viestintäteknologioiden käyttö on myös suurelta osin tekstien ja merkkien lukemista ja tuottamista. Tutkimukset osoittavat perinteisten tekstiympäristöjen ja julkaisukanavien menettävän yleisöä suhteessa digitaalisiin medioihin. Näiden seikkojen vuoksi huolipuhe läpäisee ja värittää lukutaitokeskustelua. Yhä enemmän tutkitaan ja keskustellaan siitä, miten lukutaitoa voidaan parhaiten edistää. Lukutaitotyön tueksi tarvitaan kuitenkin myös käsitetyötä siitä, mitä ylipäätään ymmärrämme ja rajaamme lukemiseksi. Mitkä ovat parhaita lähestymistapoja lukemisen edistämiseen esimerkiksi kirjastoissa ja kouluissa?

Tieteellisestä näkökulmasta keskeinen kysymys on se, millainen teoriaperusta parhaiten tukisi opetuksen ja oppimisen kehittämistä, sillä lukemista ja lukutaitoja voidaan tarkastella monista teoriaperinteistä käsin. Yksi vahva tutkimusperinne on lukutaitoja standardoidusti mittaava traditio, jonka puitteissa voidaan luoda pitkäjänteistä ja kansainvälistä vertailuaineistoa. Lukutaitokäsitteessä ja -ajattelussa on kuitenkin tapahtunut jatkuvaa kehitystä ja siirtymiä 2000-luvulla. Nykyisin monilukutaitoteoria on omaksuttu toiminnan ja opetuksen perustaksi kouluissa ja kirjastoissa.¹

Informaatiotutkimuksen piirissä tutkimusta on tehty huomattavasti enemmän informaatiolukutaidosta (*information literacy*) kuin lukutaidoista monikossa (*literacies*). Informaatiotutkimuksessa on kuitenkin aina ollut vahvana läsnä myös sosiologinen ja kulttuurisosiologinen tutkimusperinne, jossa lukijat ja lukemiseen liittyvät käsitykset ovat olleet keskeisiä kiinnostuksen kohteita.² Monilukutaitoteorian lukutaitokäsite ja laaja tekstikäsite ovat osaltaan laajentaneet myös informaatiotutkimuksen tutkimuksen fokusta tiedonhankinnasta ja -hausta (tieteen) alalukemiseen ja kirjoittamisen eli tiedontuottamisen käytäntöihin.

Lukutaitojen ja lukemisen tutkimusta tehdään monien tieteenalojen piirissä. Tässä artikkelissa jätän kuitenkin lukemisen tutkimuksen monitieteisyyden käsittelyn ulkopuolelle. Ihmis- ja yhteiskuntatieteiden luonteeseen kuuluu, että samat teoreettiset siirtymät tapahtuvat yleensä kaikilla erityistieteenaloilla, joskin ehkä eriaikaisesti. Kullakin lukemisesta ja lukutaidoista kiinnostuneella tieteenalalla on luonnollisesti myös tieteenalan perinteeseen nivoutuvia erityisiä huolenaiheita ja tulokulmia suhteessa lukemiseen.

Kirjallisuudentutkimus on perinteisesti ollut kiinnostuneempi kirjallisuuden vaikutuksesta lukijaan kuin todellisista lukijoista, joskin tämä on selvästi muuttumassa. Psykologia ja neurotieteellinen

1 <https://www.oph.fi/fi/koulutus-ja-tutkinnot/perusopetuksen-opetussuunnitelman-ydinasiat>, <https://www.oph.fi/fi/tilastot-ja-julkaisut/julkaisut/kansallinen-lukutaitostrategia-2030>

2 Esim. Ross 1987; Mäkinen 1997; Vakkari 1983.

tutkimus ovat kiinnostuneita siitä, miten lukeminen vaikuttaa erilaisiin aivoalueisiin ja miten lukemisaktiivisuus mahdollisesti näkyy aivokuvantamisessa. Vuorovaikutteinen teknologia on kiinnostunut esimerkiksi e-kirjojen ja e-kirjapalveluiden käytettävyydestä ja kehittämisestä. Lukemiskäytäntöjen tutkimus puolestaan kiinnostaa informaatiotutkimuksen lisäksi esimerkiksi kasvatustiedettä, kulttuurintutkimusta, sosiologiaa ja antropologiaa. Näiden tieteenalojen näkökulmat ja lähestymistavat eivät välttämättä juurikaan eroa toisistaan.

Lukutaidon määritelmät

Yksinkertaisimman määritelmän mukaan lukutaitoinen osaa lukea ja kirjoittaa vähintäänkin yksinkertaisia lauseita arkipäivän asioista sekä ymmärtää lukemaansa tekstiä. Perinteiseen lukutaitokäsitteeseen liitetään usein taitotason arviointi (osaa lukea auttavasti – osaa lukea sujuvasti). Esimerkiksi PISA-testit mittaavat kykyä ymmärtää kirjoitettua ja painettua tekstiä. Lukutaito ei kuitenkaan ole vain teknistä tai mekaanista lukutaitoa, sillä tekstin ymmärtäminen edellyttää myös kykyä tulkita, arvioida ja tuottaa tekstejä. Lukutaitoisuuden määritelmään liitetään yleisesti myös muita kompetensseja kuten kyky elinikäiseen oppimiseen ja aktiiviseen toimintaan yhteiskunnassa sen täysivaltaisena jäsenenä. Unesco määrittelee lukutaidon seuraavasti:

”Lukutaito on kyky tunnistaa, ymmärtää, tulkita, luoda, viestiä sekä eritellä painettuja ja kirjoitettuja tekstejä erilaissa konteksteissa. Lukutaidon kehittyminen on jatkuva oppimisprosessi, jonka avulla yksilö voi saavuttaa tavoitteitaan, kehittää tietojaan ja mahdollisuuksiaan sekä osallistua täysipainoisesti yhteisölliseen ja yhteiskunnalliseen toimintaan.”³

Unescon lukutaitomääritelmä korostaa sitä, että lukutaito on kehittyvä prosessi, joka mahdollistaa sen, että yksilö voi asettaa ja saavuttaa oppimistavoitteita ja hyödyntää täydesti mahdollisuuksiaan. Unescon määritelmä laajentaa lukutaidon kykyyn käyttää tieto- ja viestintätekniologioita. Lukutaidon ja tietoteknologisen osaamisen (*ICT literacy*) yhteenkietoutuminen on digitaalisten informaatioteknologioiden ja viestintäympäristöjen kehityksen myötä luonnollista ja itsestään selvää. Unescon määritelmä rajaa kuitenkin lukutaidon hieman vanhakantaisesti kirjoitettujen ja painettujen tekstien (*”printed and written materials”*) lukemiseen. Viestinnän, oppimisen ja tiedonmuodostuksen kontekstit ovat moninaistuneet ja digitalisoitumisen myötä lukutaitoon kuuluu kyky hyödyntää digitaalisia tietoresursseja.

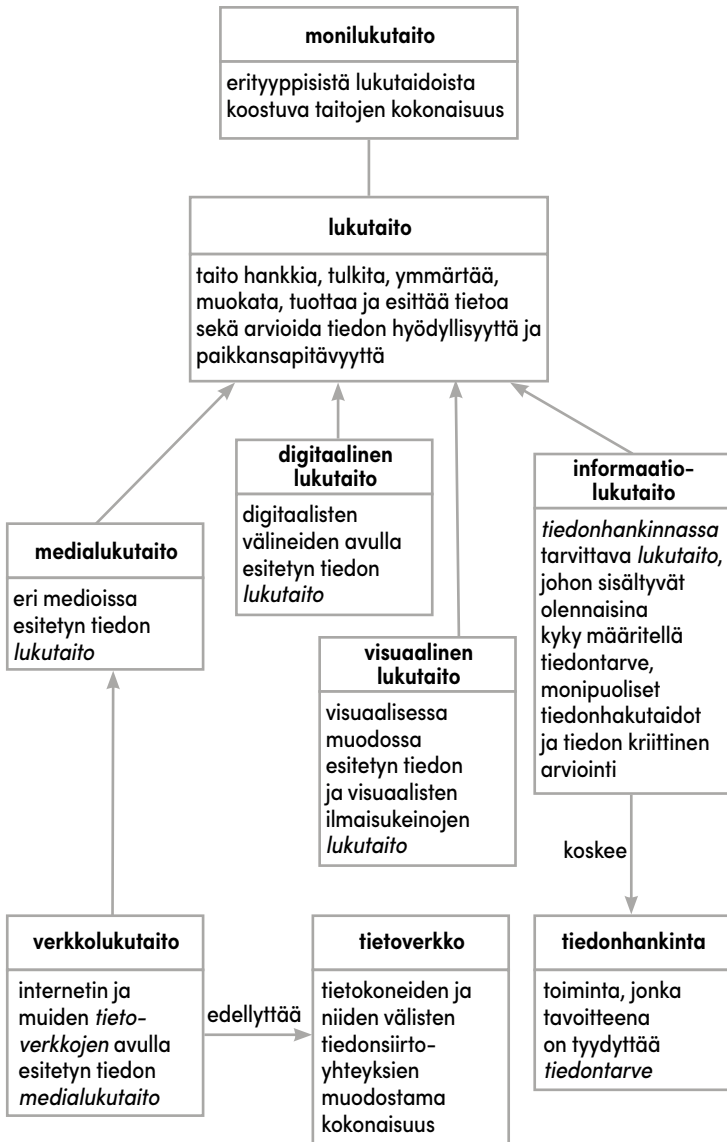
Lukutaito ei siis ole vain yksi taito, vaan rypäs erilaisia taitoja. Tämä on tavallaan itsestään selvää, kun ajatellaan lukemisen ja kirjoittamisen yhteenkietoumista viestinnän mahdollistajina. Suomalaisen asiasanasto- ja ontologiapalvelu Finton käsiteanalyysin mukaan lukutaidon osa-alueet jäsenyivät kuviossa 1 esitetyn mukaisesti.⁴

Finton käsiteanalyysi on hyvä lähtökohta lukutaitokäsitteen analyysiin, sillä se esittää, miten käsitteet ja niiden väliset suhteet on yleisimmin määritelty. Samalla kun lukutaito ymmärretään yk-

3 ”The ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning enabling an individual to achieve his or her goals, develop his or her knowledge or potentials, and to participate fully in the community and wider society.” <https://uis.unesco.org/en/glossary-term/literacy>

4 <https://finto.fi/tt/fi/page/t33>

2.2 Lukutaitojen ja lukemisen tutkimus



Kuvio 1. Finton lukutaitokäsitekaavio. Finton lisenssi alkuperäiselle kuvalle CC BY 4.0

sittäisen taidon sijaan taitojen ryppääksi, käsite-erottelun pohjana on medioiden erottelu. Verkkoalustojen tekstien ymmärretään vaativan omanlaisiaan lukutaitoja, (perinne)medioiden omanlaisiaan. Viestinnän, uutisoinnin ja keskustelun kanavien erottelu alustojen perusteella erilaisia lukutaitoja vaativaksi merkitsee sitä, että niin sanotulle perinнемедialle eli painetuille sanomalehdille ja valtiollisesti ylläpidetyille viestimille, kuten Yleisradiolle, annetaan etusija ei-problemaattisina tiedonlähteinä, kun taas uudempien viestintä-kontekstien katsotaan vaativan erityisiä kriittisen arvioinnin taitoja. Visuaalisen ja digitaalisen lukutaidon erottaminen kirjallisten painettujen tekstien lukutaidosta kuvastaa sekin perinteistä lukutaito-ajattelua, jonka mukaan lukutaidon perusta on nimenomaan kirjallinen ja painettu teksti.

Huomionarvoista määritelmässä on myös se, että lukutaitojen oletetaan olevan yksilön taitoja, vaikka tätä ei sanota suoraan. Monilukutaitoajattelu liittyy sosiokulttuuriseen teoriaperinteeseen, jossa lukutaitoa ei enää tarkastella vain yksilöiden kognitiivisena kykynä vaan kokonaisten yhteisöjen osallistavana ja vuorovaikutuksellisenä käytäntönä.⁵

Finton määritelmässä lukeminen näyttäytyy toimintana, jonka tavoitteena on aktiivinen oppiminen ja tiedonhankinta. Lukemisen oletetaan lähtevän liikkeelle tiedontarpeesta, ei esimerkiksi arjen rutiineista tai rauhoittumisen ja viihtymisen tarpeesta.

Finton määritelmä on luonnollisestikin vain yksi mahdollinen tulkinta käsitteistä ja niiden sisällöstä, ja se edustaa tiedonhankintapainotuksessaan informaatiotutkimuksen näkökulmaa. Esimerkiksi kasvatustieteen ja sosiologian piirissä ei juurikaan puhuta informaatiolukutaidosta lukutaidon erillisenä osa-alueena. Siinä, missä informaatiotutkimus on kiinnostunut nimenomaan tiedonhaun ja tiedonhankinnan taidoista ja tällä tavalla ottaa peruslukutaidon eli teknisen lukutaidon ikään kuin itsestäänselvyytenä – jonakin, joka yksilöillä on jo – kasvatustiede puolestaan ottaa tiedonhankinta-

5 Herkman & Vainikka 2012, 37–40; Kupiainen 2017, 206–207.

taidot itsestäänselvyytenä eli enimmäkseen jonakin, jotka yksilöllä on jo, joita ei tarvitse erikseen käsitellä tai opettaa. Informaatiolukutaito on enimmäkseen informaatiotutkimuksen piirissä tutkittu ja käytetty käsite. Mediakasvatusta ja -tutkimusta kiinnostaa medialukutaito, jolla tarkoitetaan enimmäkseen tiedon paikkansapitävyyden arviointia.⁶ Tietoyhteiskunta- ja digitalisaatiotutkimusta kiinnostaa digitaalisten taitojen käsite,⁷ joka enimmäkseen ymmärretään pääsyyntä digitaalisiin tietoresursseihin eli laitteiden ja digitaalisten taitojen levinneisyytenä globaalilla tasolla. Äidinkielen ja kirjallisuudentutkimuksen näkökulmasta lukutaidon mittarina eivät ole tiedonhaun tai arvioinnin taidot, vaan laajemmin ymmärrettynä tekstitaidot.⁸ Taidekasvatusnäkökulma kirjallisuudenopetuksessa puolestaan ensisijaistaa kokonaisten kaunokirjallisten teosten lukemisen kyvyn.

Lukutaito-käsiteperheessä näkyy, kuinka myös monilukutaidon käsite taipuu eri tulkintoihin riippuen taustateoriasta. Sosiokulttuurinen lukutaitojen tutkimus on kiinnostunut kaikesta lukemisesta, ei vain tiedonhankintaan liittyvästä lukemisesta. Lukijan ei oleteta olevan lähtökohtaisesti tiedon tarpeessa (ei-tietävä, ei-informoitu), vaan kiinnostus kohdistuu pikemminkin siihen, miten lukijan jo olemassa olevat kokemukset ja tausta vaikuttavat lukukokemukseen ja tekstin tulkintaan ja kuinka lukutaidot kehittyvät yhteisöllisesti ja teknologioiden kehityksen myötä.⁹ Tämän vuoksi lukutaitojen sijaan puhutaan mieluummin lukutaitokäytännöistä.¹⁰ Lukutaitokäytäntöjen kehityksessä nähdään olevan kyse sosiokulttuuristen ja teknologismateriaalisten käytänteiden muutoksista sekä historiallisista siirtymistä sen suhteen, kenen ja missä esitetty tieto lasketaan autoritaariseksi tai relevantiksi.

6 <https://medialukutaitosuomessa.fi/mediakasvatuslinjaukset.pdf>

7 <https://signals.fi/kategoria/digitaalisuus/hyva-saavutettavuus-tukee-digitaalisen-lukutaidon-kehittymista/#3b70427d>

8 Grünthal 2020.

9 Tuominen ym. 2005.

10 Street 2003; Hoffman ym. 2016; Käytäntöteoriasta ja käytännön käsitteestä, ks. Talja 2010.

Informaatiolukutaito, digitaalinen lukutaito ja lukutaitokäytännöt

Informaatiolukutaidon, digitaalisen lukutaidon ja monilukutaidon käsitteet saavat erilaisia tulkintoja riippuen tulkitsijan tausta-teoriasta. Esimerkkinä tästä ovat muutokset informaatiolukutaidon ja digitaalisten lukutaitojen määritelmissä.

Yhden määritelmän mukaan digitaalinen lukutaito mittaa ihmisten kykyä löytää relevantteja (tiedontarpeisiin vastaavia) sisältöjä verkosta.¹¹ Digitaaliset taidot näyttäytyvät mitattavina.¹² Taitoja mitataan ensi sijassa usein yleisimpien sovellusten käytön kautta. Myös tietokonelukutaidon tai tietoteknisen lukutaidon mittarina on käytetty yleisimpien ohjelmien käyttötaitoa. Taustaoletuksena on ohjelmien geneerisyys: oletetaan, että kaikilla ihmisillä on samanlainen tarve hallita tietyt perussovellukset ja -ohjelmat (tietokoneajokortit edustavat samaa ajattelutapaa).¹³ Näkökulmana ei tällöin ole ihmisten toiminta työn ja arjen konteksteissa, vaan ohjelmien tarjoumat ja toiminnallisuudet oletetaan yleisesti hyödyllisiksi ja tarpeellisiksi työtehtävistä riippumatta. Taitoja on yleisesti mitattu kyselytutkimuksin, joissa kysymykset koskevat erilaisten sovellusten käyttöä sekä vastaajien arvioita omista taidoistaan (minäpystyvyysoletukset,¹⁴ *self-efficacy*). Motivaation lukutaitojen hankintaan ja laajentamiseen on katsottu olevan riippuvaista minäpystyvyysoletuksista eli tietoteknisestä minäkuvasta tai lukijaminäkuvasta.¹⁵ Tämä on luonnollisestikin kehäpäätelmä.

Monilukutaitoteorian¹⁶ kehittämisen yhtenä lähtökohtana on ollut halu päästä irti sovellusten, medioiden, ohjelmien tai kirjallisuuden olemassaolon asettamisesta taitojen ja kyvykkyyksien

11 Ks. Hargittai 2002.

12 Esim. Li & Hu 2022.

13 Talja 2005.

14 Bandura 1986.

15 Talja 2005.

16 New London Group 1996.

mittariksi. Lukutaitostandardien eli testaus- ja mittaamistradition sijaan digitaalisia taitoja voidaan lähestyä siitä näkökulmasta, mitä digitaaliset alustat eli tieto- ja viestintäympäristöt mahdollistavat verrattuna perinnemediaan. Gilster¹⁷ määrittelee digitaalisten tietojen ja viestiympäristöjen mahdollisuudet eli digitaaliset lukutaitokäytännöt seuraavasti:

- laajentunut tietoisuus muiden ihmisten ajatuksista, maailmoista ja tunteista
- laajentunut pääsy informaatioon, tietoon ja oppimiseen perinteisen koulutuksen ulkopuolella
- tiedollisten mahdollisuuksien laajentuminen koulutukseen valittujen tai muuten etuoikeutettujen ryhmien ulkopuolelle.

Lankshear ja Knobel loivat termin uuslukutaidot (*new literacies*) korostaen sitä, että uudet teknologiat vaikuttavat dynaamisesti ja vuorovaikutteisesti uudenlaisten viestintäympäristöjen kehittämiseen.¹⁸ Nämä ympäristöt sisältävät erilaisia arvostuksia ja toimintamalleja kuin konventionaaliset lukutaitoympäristöt. Uudet lukutaitokäytännöt puolestaan muovaavat sitä, millaisiksi teknologiset alustat ja niiden tarjoumat kehittyvät. Lankshearin ja Knobelin mukaan uusille lukutaitokäytännöille on ominaista toimijoiden sitoutuminen yhteistyöhön, osallisuus tiedon ja kokemusten tuotantoon, tuen ja neuvojen vapaa jakaminen, yhteisen edun korostus sekä kilpailullisuuden poissaolo. Uuslukutaitokäytännöt antavat toimijoille sisäpuolisen osallistujan roolin, ja niiden ydintä on tiedon ja kokemusten jakaminen vertaistalossa.¹⁹

Monilukutaitoteoria on kiinnostunut siitä, miten oppijoiden omassa arjessaan omaksumat lukutaitokäytännöt voidaan sovittaa

17 Gilster 1997.

18 Lankshear & Knobel 2007.

19 Lankshear & Knobel 2007.

yhteen virallisen kouluopetuksen kanssa.²⁰ Kouluissa oppija pyrkii enimmäkseen toteuttamaan ulkopuolelta, opetussuunnitelmissa annettuja tavoitteita, esimerkiksi lukemaan ja omaksumaan tietyn määrän ja tietynlaisia tekstejä (oppikirjatekstejä ja kaunokirjallisuutta). Oppija on suoriutumisen arvioinnin kohteena, vaikka hänelle annettaisiin vapauksia tekstien ja kiinnostuksen kohteiden valinnassa. Asetelma on siis kilpailullinen, eikä oppijan roolina ole tuottaa mitään, mikä hyödyttäisi, tukisi tai ilahduttaisi myös muita ihmisiä samassa vertais- tai intressiyhteisössä.

Kun motivaatiota ja minäpystyvyyksäilyksiä pidetään oleellisina suoriutumisen eli koulun arvostamien lukutaitokäytäntöjen kehittymisen kannalta, yhtälö on ymmärrettävästikin hankala. Uutisotsikot esimerkiksi PISA-testien heikkenevistä tuloksista tai poikien tyttöjä heikommasta suoriutumisesta PISA-testeissä porautuvat harvemmin pintaa syvemmälle. Monilukutaitoajattelu korostaa sitä, että on tärkeää kysyä, mikä on toimijan oma rooli tiedon tuottajana ja tiedon käyttäjänä. Kuinka koulu pystyy tunnistamaan ja huomioimaan tiedon ja toiminnan aloja ja alueita, joissa oppija on asiantuntija eikä aloittelija? Onko aina oikein lähteä liikkeelle oletuksesta, että toimija on lähtökohtaisesti tyhjä astia, joka täytetään tiedolla ja kriittisellä arviointikyvyllä?

Perinteiset informaatio- ja medialukutaitokäsitteet korostavat kriittisyyttä ja luettelevat erilaisia kriteerejä ja tunnusmerkkejä luotettaville tiedonlähteille.²¹ Uuslukutaitokäsite ja -ajattelu poikkeaa perinteisestä informaatiolukutaitokäsitteestä, joka olettaa toimijan etsivän ja koostavan tietoa ulkopuolisista, parhaassa tapauksessa virallisista, vakiintuneista ja arvovaltaisista lähteistä sekä suhtautuvan kriittisesti esimerkiksi vertaisyhteisöihin. Kriittistä informaatio- ja medialukutaitoa korostavien tyyppillisen väittämän mukaan ”verkko on täynnä vihapuhetta ja valetietoa”.²² Tällöin voi olettaa, että puhujan intressinä on perinnemedioiden heikkenevän arvovallan

20 Cope ym. 2018, 6–7; Kupiainen 2021, 76.

21 Tuominen ym. 2005.

22 Ks. esim. <https://medialukutaitosuomessa.fi/mediakasvatuslinjaukset.pdf>

ja erityisaseman puolustaminen ja että puhuja ei juurikaan liiku verkon erilaisilla alustoilla, joissa toimijat yhdessä jakavat, tuottavat ja kehittävät alansa tietoa ja kieltä jaettujen intressiensä pohjalta.

Tiedonhaku on arjessa saumaton osa verkon eli tietokoneen ja kännykän käyttöä samoin kuin lukeminen.²³ Tiedonhaku ja tiedonhankinta olivat erityinen ja erityisiä taitoja vaativa toiminto silloin, kun tieteelliset ja muut artikkelit oli haettava bibliografisista tietokannoista ja noudettava kirjastoista. Informaatiolukutaito-termiä edelsikin termi kirjastonkäyttökoulutus.²⁴ Sosiokulttuurisen näkökulman läpimurron myötä informaatiolukutaidon teoria ja tutkimus erkanivat tiedonhankinnan opetuksesta ja informaatiolukutaitostandardeista.²⁵ Tässä perinteessä informaatiolukutaitotutkijoiden kiinnostus kohdistuu ennen muuta siihen, kuinka ymmärtäminen, asiantuntemus ja osaaminen kehittyvät ja ilmenevät yhteisöllisissä käytännöissä ja toiminnoissa eri konteksteissa.²⁶

Perinteisen informaatiolukutaitoajattelun puitteissa oppijoita tuli ohjata virallisen oppikirjatotuuden ja autoritääristen lähteiden äärelle. Kollektiivisella tasolla tieto ja tiede ovat kuitenkin dialogisia, jatkuvasti kehittyviä prosesseja. Tiede suorittaa jatkuvia korjausliikkeitä ja muuttaa kysymyksenasettelujaan, jos ne uusien löydösten valossa alkavatkin näyttää huonosti tai väärin asetelluilta. Esimerkiksi informaatiolukutaitotutkimuksen on väitetty rakentaneen ja määritelleen ongelman (kyvyttömyys hakea ja arvioida tietoa), joka ei ole tosiasiallisesti olemassa oleva ongelma.²⁷ Tämän tyyppisiä tieteellisiä kiistoja, ristiriitoja ja teoreettisia käänteitä pidetään tieteen puitteissa hedelmällisinä tiedon ja ymmärryksen edistymisen kannalta. Tältä kannalta ajateltuna yksi lukutaitokäytäntöjen kehittymisen edellytys on liikkuminen monissa erilaisissa keskusteluyh-

23 Haider & Sundin 2019.

24 Tuominen ym. 2005.

25 Hicks 2018.

26 Limberg ym. 2012.

27 Tuominen ym. 2005.

teisöissä ja medioissa sekä kiinnostuksen ja huolenaiheiden jakaminen muiden kanssa niin tiedon vastaanottajana kuin sen tuottajana.

Mitä lukeminen on?

Sosiokulttuurisesta näkökulmasta lukutaito voidaan määritellä prosessiksi, jossa lukemalla, kirjoittamalla ja puhumalla tuotetaan merkityksiä teksteistä sosiaalisesti paikantuneissa käytännöissä vuorovaikutuksen ja osallistumisen kautta.²⁸ Määritelmä ottaa lähtökohdakseen viestintäympäristöjen ja -kontekstien moninaisuuden ja vaihtelevuuden ensisijaistamatta virallisia kielenkäyttöympäristöjä (oppikirja, sanomalehti tai tv-uutiset).

Arkielämässä toimijat osaavat luontevasti liikkua vaihtelevissa viestinnällisissä ja kielellisissä tilanteissa ja ottaa niissä erilaisia toimijarooleja. Yhdessä kontekstissa, vaikkapa kurssille mennessään, sama henkilö on oppija ja sovittaa toimintansa tähän kontekstiin sopivaksi. Toisessa kontekstissa, vaikkapa opastaessaan työpaikalla uutta työntekijää, sama henkilö on opettaja. Yhdessä kontekstissa toimijuutta määrittää velvollisuus kuunnella, toisessa velvollisuus esittää selkeästi talon tavat. Samat viestinnälliset kyvykkyydet, jotka pätevät arkielämässä tietoverkkojen ulkopuolella pätevät luonnollisesti myös erilaisilla teknologisilla alustoilla, joissa jokaisessa on omat keskustelu- ja tiedontuottamistraditionsa. Liikkuminen tieto- ja viestintäympäristöjen välillä ja osallistuminen erilaisiin vertais-yhteisöihin ja niiden käytäntöihin on lukutaitojen kehittymisen ydintä.

Lukeminen määritellään merkitysten tuottamiseksi teksteistä, joko lukemalla tai kuuntelemalla.²⁹ Vieläkin laajemmin lukemisen voi määritellä merkitysten tulkitsemiseksi ja ymmärtämiseksi erilaisissa konteksteissa kuten katsottaessa elokuvaa, pelattaessa videopeliä tai kuunneltaessa podcastia.³⁰ Suuri osa lukemisesta tapah-

28 Frankel ym. 2016, 7.

29 Tattersall Wallin & Nolin 2020, 471.

30 Tattersall Wallin & Nolin 2020, 471.

tuu nykyisin verkossa, esimerkkinä sosiaalisen median päivitysten, kommenttien ja uutisotsikoiden lukeminen. Miksi lukeminen tai tekstit ovat mielekkäitä ymmärtää näin laajasti? Esimerkiksi äänikirjojen kuuntelu on jatkumo suullisen tarinankerronnan traditiolle, joka edeltää painettuja tekstejä. Kirjojen osalta lukeminen ääneen ja lukeminen muille on keskeinen osa lukemisen historiaa. Äänikirjat auttavat lukemisen tilojen ja paikkojen laajentamisessa.

Laaja tekstikäsitelmä korostaa merkitysten luomista lukemisen ytimenä. Myös musiikin kuuntelu on tulkintaa ja varsinkin lyriikoiden osalta runouden tulkintaa. Laululyriikat ovat ikaikaaisesti olleet keskeisellä sijalla kansan muistissa ja tajunnassa samalla tavoin kuin tarinat, kummitusjutut, lorut ja sananlaskut. Tv-sarjojen katselija lukee tekstityksiä ja kuuntelee kieltä sekä sen ääniteitä ja vivahteita. Tämä monikielinen lukutaitokäytäntö ei ole olemassa niissä maissa, joissa tv-sarjat ovat dubattuja, mutta merkitysten luominen ja tulkinta, eläytyminen sekä viestinnällisyys (josta voidaan puhua myös tietona), ovat tässäkin läsnä.

Kiinnostavaa on se, millaisin strategioin toimijat luovat, irrottavat ja purkavat viestejä ja merkityksiä erilaisissa tekstiympäristöissä. Verkkotekstejä luetaan monesti silmäillen ja poimien nopeasti kiinnostavat kohdat tekstimassasta. Digitaalisen lukemisen oletetaan monesti johtavan pinnalliseen lukemiseen ja keskittymiskyvyn heikkenemiseen.³¹ Toisaalta kyse on myös siitä, että seulova ja poimiva lukustrategia on mielekäs monessa lukemisympäristössä. Lukija syventyy siihen, mikä häntä itseään kiinnostaa. Monimediallisessa ympäristössä tulkinta ja merkitysten luominen kohdistuu tekstin ohella kuvalliseen ilmaisuun, ääniin ja moniin teknisiin seikkoihin. Suuri osa tällaisesta lukemisesta ja tulkinnasta tapahtuu rutiininomaisesti ja ilman tietoista ponnistelua, mikä ei vähennä sen tärkeyttä tai merkitystä.

31 Wolf 2019.

Lukemisen tutkimuksen tärkeä huomio on, että suuri osa lukemisesta on huomaamatonta.³² Tutkimuksissa on havaittu, että moni lehtien, tietokirjallisuuden tai verkkotekstien lukemista aktiivisesti päivittäin harrastava ei määrittele itseään ”lukijaksi”. ”Lukijuus” ymmärretään esimerkiksi omassa kulttuurissamme erityiseksi identiteetiksi, kaunokirjallisuuden jatkuvaksi ja aktiiviseksi lukemisharrastukseksi. Yksi syy siihen, että lukeminen näyttää tilastollisesti tarkastellen vähentyvän, on suppea määritelmä lukijasta ja lukemisesta. Emme vain tule ajatelleeksi sitä, miten suuressa määrin arkinen lukeminen jakaantuu ja hajautuu erilaisille alustoille ja ympäristöihin. Halu puolustaa ja edistää kaunokirjallisuuden ja varsinkin vakavan kaunokirjallisuuden asemaa saa monesti vähättelemään esimerkiksi äänikirjojen kuuntelua oikean lukemisen korvikkeena. Monesti tilanne on vain se, että monen tietotyöläisen työ- ja arkielämä edellyttää päivittäin niin paljon erilaisten ja usein vaativienkin tekstien lukemista, että aivokapasiteetti ja aikabudjetti eivät enää helposti veny kiinnostavan kaunokirjallisuuden lukemiseen. Podcastit ja äänikirjat vapauttavat kehon ja kädet muihin toimiin samalla, kun toimija lukee kuuntelemalla.

Empatian, mielikuvituksen, sanavaraston ja ilmaisukykyjen kehittyminen liitetään kulttuurissamme tyypillisesti kaunokirjallisuuden lukemiseen ja perinteisesti erityisesti niin sanottuun laatukirjallisuuteen. Toisessa kulttuurissa vetäytymistä yksin kirjallisuuden pariin eli hiljaista yksinlukemista voidaan pitää epäsosiaalisena tai epänormaalina.³³

Perinteisesti lukeminen on ymmärretty nimenomaan erilliseksi yksittäiseksi toiminnoksi.³⁴ Lukemisen myönteisten vaikutusten on osaltaan nähty seuraavan siitä, että lukeminen edellyttää monimutkaisia kognitiivisia prosesseja. Lukemisen prosessi ei kuitenkaan ole vain yksilön mielen sisäinen vaan myös ruumiillinen toiminto, joka

32 Asplund & Perez Prieto 2018; Harju 2018.

33 Smagorinsky 2009.

34 Smagorinsky 2009.

vaatii silmien, aivojen ja käsien koordinoitua.³⁵ Lukija osallistuu sosiokulttuuriseen käytäntöön eli oppii muodostamaan nopeaa tai vaivatonta etenemistä helpottavia odotuksia tekstien suhteen, esimerkiksi genresääntöjä. Hyväksi lukijaksi kehittyminen vaatii luonnollisestikin paljon harjoittelua ja toistoa kuten minkä tahansa käytännön, esimerkiksi jääkiekon pelaamisen tai musiikki-instrumentin soittamisen, oppiminen.³⁶

Lukemisen tärkeyttä korostavat ensisijaistavat monesti nimenomaan tekstien tulkinnan taidon eli metakognition – lukemisen hetkellisestä ajallisesta kokemuksesta irrotetun jälkikäteisanalyysin. Kaikesta huolimatta lukijan tulkinta on riippuvainen hänen omista kokemuksistaan – sekä aiemmista lukukokemuksista että yleisistä elämäkokemuksista ja -piiristä. Lukeminen voi antaa tilan ja paikan kokea ja kuvitella uudenlaisia minuuksia ja maailmoja, mutta toisaalta samanlaisia oivalluksia ja tunteita on mahdollista saada muistakin medioista, muun muassa elokuvista ja tv-sarjoista. Tekstin sisältöjen ja vaikuttamiskeinojen jälkikäteisanalyysi ja -kriittikki tai oman oppimisen tarkastelu ovat samalla tavoin mahdollista näissä muissa konteksteissa. Usein lukemisen erityisiä vaikutuksia ja etuja korostavat ovat kirjallisuuden, eivät niinkään lukemisen, edistämisen asialla.³⁷ Tämän vuoksi lukemisen tutkimus eroaa kirjallisuuden tutkimuksesta.

Lukutaitokäytäntöjen alasisidonnaisuus

Merkityksen muodostaminen tekstistä tapahtuu aina jossakin kontekstissa. Kokeminen ja kokemuksen artikulointi eli sanallistaminen ovat kaksi eri asiaa.³⁸ Tilanteinen lukukokemus elää hetkessä, kun taas kertomus tai yhteenveto luetusta, tulkinta, eroaa siitä ajallisesti ja on suunnattu jollekulle ja jotakin tarkoitusta varten – joko itselle

35 Trasmundi ym. 2021.

36 Ross 2009, 651.

37 Hornby 2006, 6–8.

38 Balling 2016.

tai toiselle. Persson³⁹ korostaa, että kirjallisuudentutkijoiden ja kriitikoiden suosima tekstianalyttinen lukutapa, jota usein pidetään elämyskokemuksellista lukutapaa kehittyneempänä, on sekin vain yksi sosiokulttuurinen lukutaitokäytäntö. Kriittisen lukutavan ajatellaan aina suuntautuvan tietyllä tapaa: purkamaan sitä, kuinka teksti pyrkii vaikuttamaan erilaisilla retorisilla ja tyylikeinoilla, ja punnitsemaan sitä, millaista ideologiaa se puoltaa ja millaista maailmaa se luo. Kriittinen analyysi edellyttää omalla tavallaan tekstin kanssa vastahankaan asettumista, kun taas ”tavallisen” lukemisen oletetaan olevan enemmän henkilökohtaista, tunnepohjaista ja myötäsukaista.⁴⁰ Lienee selvää, ettei lukutapaa voida irrottaa siitä genrestä ja traditiosta, jonka sisällä luetaan. Kriittisen ja epäilysten vallassa olevan lukijan lukukokemus ei ole laadukkaampi tai arvokkaampi kuin nau-ravan tai itkevän tunteellisen lukijan.

Sosiokulttuurinen näkökulma korostaa lukutaitokäytäntöjen sosiaalisuutta ja paikantuneisuutta. Koulu on omanlaisensa sosiokulttuurinen konteksti, jossa luetaan, kirjoitetaan ja viestitään tietyllä tapaa. Akateeminen tieteenala on vastaavanlaisesti omanlaisensa lukutaitokäytäntö, jonka puitteissa opiskelijat ja tutkijat oppivat kirjoittamaan tietyllä tyylillä ja välttämään erityisiä huolenaiheita ja arvoja.⁴¹ Tutkijoiden tulee hallita alansa teorioiden, historian ja menetelmien ohella monenlaisia kulttuurisia jaettuja valmiuksia. Tavat ja tyylit, joilla tuloksia esitetään ja ongelmia asetetaan ja joilla lukija pyritään vakuuttamaan, eroavat eri tieteenaloilla toisistaan.⁴² Tämän vuoksi esimerkiksi sosiologi ei välttämättä pysty lukemaan fysiikan alan artikkelia. Vastaavasti fysiikan alan tutkija ei välttämättä tule ajatelleeksi, että kirjallisuudentutkijalle esteettinen ja luova kielenkäyttö voi olla yhtä tärkeää kuin yksittäiset tulokset tai argumentit.

39 Persson 2016.

40 Persson 2016.

41 Moje 2015, 257–258.

42 Talja ym. 2007.

Yksittäinen tutkija voi olla yksin oman artikkelinsa tekijä ja kirjoittaja, mutta tosiasiallisesti taustalla on kokonainen joukkue eli satoja ja tuhansia aiempia tutkijoita, jotka ovat osaltaan yhdessä rajanneet ja puitteistaneet kentän, jolla pelataan.⁴³ Kukin tieteenala on oma lajinsa, kuten keilaaminen tai jääkiekko. Kuhunkin lajiin liittyy myös kulttuurisia vuorovaikutuksen ja viestinnän ”pukuhuonetraditioita” eli puheen ja puhuttelun sääntöjä, jotka aloittelija vähitellen omaksuu lajin varsinaisten taitojen ja keinojen ohessa.⁴⁴ Lukutaitokäytäntöjen vuorovaikutteisen kehityksen ja vähittäisen oppimisen sääntö koskee luonnollisesti kaikkia moninaisia työ- ja harrastusaloja kuten käsitöitä ja autonkorjausta. Käsityöharrastajan on vaikeaa lukea autonkorjausoppaita ja päinvastoin. Osaamisalueiden erojen tunnustaminen on keskeinen lukemisen tutkimuksen lähtökohta. Kiinnostavia tuloksia voidaan saada pyrkimällä hienovaraisesti havainnoimaan alan osaajien toimintaa, vuorovaikutuskäytäntöjä sekä tiedonmuodostuksen tiloja ja paikkoja.

Sosiokulttuurisen teoriaperinteen kehittymisen myötä informaatiolukutaitoa koskeva ajattelu on muuttunut. Samalla kirjastomaailmassa on siirrytty informaatiolukutaitostandeista teoreettisesti orientoituneisiin lukutaitokehyksiin.⁴⁵ Association of College and Research Libraries (ACRL)⁴⁶ on rakentanut informaatiolukutaitokehykset, jotka sopivat hyvin silloittamaan eri aloilla tehtävää lukutaitokäytäntöjen tutkimusta ja pedagogisia käytäntöjä. Tämän artikkelin kontekstiin soveltaen lukutaitokehyksiä eli teoreettisia lähtökohtia ovat:

- Tekstien auktoriteetti on kontekstuaalista, eli medioita tai tiedonlähteitä ei voida jakaa yksiselitteisesti luotettaviin tai epäluotettaviin, tiedollisiin tai ei-tiedollisiin.

43 Cronin 2004.

44 Becher & Trowler 2001.

45 Hicks 2018.

46 <http://www.ala.org/acrl/standards/ilframework>

- Merkitysten luominen on prosessi. Tekstit eivät sinällään sisällä informaatiota tai sisältöjä, jotka lukija voisi poimia niistä tekemättä samalla omaa huomaamatonta työtään tekstin tulkitsijana. Informaation irrottaminen tekstistä eli merkityksen luominen riippuu lukijan taustasta ja kaikesta muusta aiemmin luetusta.
- Informaatio pohjautuu arvoihin. Informaation luomiseksi on ensin oltava ongelmanasettelu tai huolenaihe. Mikäli lukija ei jaa samaa kiinnostuksen tai huolen aihetta, informaatio ei näydy hänelle arvokkaana informaationa.
- Tieto ja tutkimus on jatkuvasti kehittyvää ja muuttuvaa (avoin työprosessi). Informaation siirtäminen ja välittäminen ”paketteina” oppikirja- ja ensyklopediatyyliin ei vastaa aitoa tutkimusprosessia, jossa aiempien tulosten ja evidenssin haastamisella ja testaamisella on aina keskeinen sija.
- Tiede ja tutkimus on keskustelua.
- Tiedonhankinta ja -haku on strategista löytöretkeilyä. Tiedontarpeet ja -intressit muuttuvat sekä yksilöllisesti että yhteisöllisesti saadun tiedon ja uusien tulosten perusteella.

Informaatiolukutaitoteorioissa puhutaan informaatiosta, kun taas lukutaitokeskustelu ja lukemisen tutkimus puhuvat laajemmin teksteistä. Kuitenkin myös informaatiotutkimuksen määritelmä informaatiosta on väljä: termi ei viittaa faktoihin tai virallisiin viesteihin vaan merkitysten tuottamiseen.

Yhteenveto

On hieman paradoksaalista, että informaatio- ja viestintäteknologioiden ja älypuhelinien muuttuessa osaksi arkielämää tutkimukset ihmisten arkisesta vuorovaikutuksesta erilaisten tekstien ja sisältöjen kanssa eivät ole lisääntyneet samaa tahtia. Lukemista tarkas-

tellaan tyypillisesti huolipuheen kautta.⁴⁷ Suomen koululaisten ja erityisesti poikien heikkenivät tulokset PISA-testeissä ovat vuosittain otsikoissa. Mäkinen⁴⁸ huomauttaa, että kun historiallisesti huolehdiin siitä, että ihmiset (naiset) lukevat liikaa ja etenkin liian ala-arvoista kirjallisuutta, nyt kaikki lukeminen saa positiivisen arvolatauksen. Lukeminen ymmärretään huolipuheessa kuitenkin yleensä kaunokirjallisuuden lukemiseksi. Aiemmin lukemista mitattiin yleisesti väestötasolla vapaa-ajan käyttötilastoissa kysymällä, kuinka monta kirjaa vuodessa vastaaja on lukenut – kirjalla tarkoitettiin painettua kirjaa.

Kansalaisten lukuintoa tarkastellaan usein myös kirjastonkäyttötilastojen kautta. Lainaustilastot eivät kuitenkaan tavoita esimerkiksi erilaisten lehtien tai tietokirjallisuuden lukemista, saati niitä moninaisia tiloja, joissa ihmiset liikkuvat internetissä – lukien ja kirjoittaen.

Etnografisen lukemistutkimuksen kiinnostuksen kohteena on lukeminen arkisena rutiininomaisena toimintana kaikissa erilaisissa konteksteissaan. Lukemisen tutkimus ei ole kiinnostunut vain aktiivisesta kaunokirjallisuuden lukemisesta tai painetuista tai julkaistuista teksteistä. Kaikki lukeminen arkisissa paikoissa ja tiloissa – työn vuoksi, toimistossa, kampaamossa, sairaalassa, harrastesivustolla, Facebookissa tai bussissa – on lukemisen tutkimuksen kohde. Tietynlaiset sisällöt ja formaatit eivät ole etusijalla. Lukemisen tutkimuksen ja kirjallisuudentutkimuksen ero on siinä, että tietynlaisten tekstien ja sisältöjen ensisijaistamisen sijaan lukemisen tutkimuksen kiinnostus kohdistuu lukemisen käytäntöihin.⁴⁹

Lukemiskäytäntöjen tutkimus pohjautuu sosiokulttuuriseen teoriaan, joka korostaa sitä, että mikä tahansa teksti tulee eri tavalla vastaanotetuksi ja tulkituksi riippuen lukijan aiemmista kokemuksista ja kiinnostuksen kohteista. Kulttuuritausta vertaisyyhteisöineen

47 Mäkinen 2018.

48 Mäkinen 2018.

49 Rosen 2015.

ja arvoineen vaikuttaa lukemisen käytäntöihin, samoin se tila ja paikka, jossa lukija kulloinkin on – fyysisesti ja henkisesti.

Koulu- ja opiskelukontekstissa tapahtuva lukeminen on erityistapausta siinä, että lukemiseen, lukemisen menetelmiin ja lukemismotivaatioon pyritään vaikuttamaan erilaisin keinoin. Muut lukemisen kontekstit ovat enimmäkseen irrallaan suorittamisen vaateista ja vapauden valtakuntaa. Mittarina lukemisen tasolle ja laadulle toimii ainoastaan lukijan oma kokemus. Kokemus ei luonnollisestikaan aina välttämättä tarkoita positiivista kokemusta – välinpitämättömyys, tylsyys ja ärtyminen ovat lukemisen kokemuksia siinä missä ilahduminen, uppoutuminen, samastuminen ja oivalluskin. Teksti, joka on lukijaa kiinnostavasta aiheesta, voi olla sekä informatiivinen että viihdyttävä samalla kertaa.

Lukemisella on lukuisia etuja, sillä se kehittää viestintä- ja ilmaisutaitoja sekä laajentaa tietämystä ja sanavarastoa. Ei kuitenkaan voida olettaa, että taitava lukijuus yhdellä alueella siirtyy ongelmitta kontekstista toiseen.⁵⁰ Lukemista edistettäessä oleellista on lähteä liikkeelle toimijan omista kiinnostuksen kohteista. Lähestymistapaa, jossa kirjat ja lukeminen assosioituvat työnteeseen, ponnisteluun, velvollisuuksiin, suoriutumiseen tai koulumaisuuteen, kutsutaan lukumurhaksi (*readicide*).⁵¹ Kirjastomaailmassa on yleisesti lähdetty lukemisen edistämiseen korostamalla lukemisen tai kirjojen hauskuutta, nautinnollisuutta ja kiehtovuutta. Lukemisella sinällään ei ole itseisarvoa, vaan arvo syntyy – jos on syntyäkseen – lukijan kokemuksena. Laadukkuus ja nautinto ovat kokonaan lukijan elämyksen laatua.

50 Talja & Lloyd 2010; Moje 2015.

51 Gallagher 2010; Ross ym. 2018.

Lähteet

- ACRL (2015) *Framework for information literacy for higher education*. Chicago: Association of College & Research Libraries. <http://www.ala.org/acrl/standards/ilframework>
- Asplund, Stig-Börje & Hector Pérez Prieto (2018) Young working-class men do not read: or do they? Challenging the dominant discourse of reading. *Gender & education* 30:8, 1048–1064. <https://doi.org/10.1080/09540253.2017.1303825>
- Balling, Gitte (2016) What is a reading experience: Development of a theoretical and practical understanding. Teoksessa Paulette M. Rothbauer & Kjell Ivar Skjerdingsstad & Lynne McKechnie & Knut Oterholm (toim.) *Plotting the reading experience: Theory/practice/politics*. Waterloo, ON: Wilfrid Laurier University Press, 37–53.
- Bandura, Albert (1986) *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Becher, Tony & Paul R. Trowler (2001) *Academic tribes and territories*. 2. painos. Lontoo: The Society for Research into Higher Education & Open University Press.
- Cope, Bill & Mary Kalantzis & Anna Smith (2018) Pedagogies and literacies, disentangling the historical threads: An interview with Bill Cope and Mary Kalantzis. *Theory into practice* 57:1, 5–11. <https://doi.org/10.1080/00405841.2017.1390332>
- Cronin, Blaise (2004) Bowling alone together: Academic writing as distributed cognition. *Journal of the American society for information science and technology* 55:6, 557–560. <https://doi.org/10.1002/asi.10406>
- Frankel, Katherine K. & Bryce L.C. Becker & Marjorie W. Rowe & P. David Pearson (2016) From “what is reading?” to what is literacy? *Journal of education* 196:3, 7–17. <https://doi.org/10.1177%2F002205741619600303>
- Gallagher, Kelly (2010) Reversing readicide. *Educational leadership* 67:6, 36–41. <https://ingvhirannar.com/wp-content/uploads/2014/09/Reversing-readicide.pdf>
- Gilster, Paul (1997) *Digital literacy*. New York: John Wiley.
- Grünthal, Satu (2020) Lukutaito ja lukeminen. Teoksessa Liisa Tainio & Maria Ahlholm & Satu Grünthal & Sirke Happonen & Riitta Juvonen & Ulla Karvonen & Sara Routarinne (toim.) *Suomen kieli ja kirjallisuus koulussa*. Ainedidaktisia tutkimuksia 18. Helsinki: Suomen ainedidaktisten tutkimusseuran julkaisuja, 165–206.
- Haider, Jutta & Olof Sundin (2019) *Invisible search and online search engines: The ubiquity of search in everyday life*. London: Taylor & Francis.
- Hargittai, Esther (2002) Second-level digital divide: Differences in people’s online skills. *First Monday* 7:4. http://firstmonday.org/issues/issue7_4/hargittai/index.html
- Harju, Julia (2018) Osaavatko teinit enää lukea? Lukemisen kulttuurit yläkouluikäisten kirjoituksissa. *Kasvatus & Aika* 12:2, 50–61. <https://journal.fi/kasvatusja aika/article/view/74130>

- Herkman, Juha & Eliisa Vainikka (2012) Lukemisen tavat: lukeminen sosiaalisen median aikakaudella. Tampere: Tampere University Press. https://trepo.tuni.fi/bitstream/handle/10024/66381/Lukemi_tavat_2012.pdf
- Hicks, Alison (2018) Making the case for a sociocultural perspective on information literacy. Teoksessa Karen P. Nicholson & Maura Seale (toim.) *The politics of theory and the practice of critical librarianship*. Sacramento: Library Juice Press, 69–85. <https://discovery.ucl.ac.uk/id/eprint/10044627/1/Hicks.pdf>
- Hoffman, James V. & Ramon A. Martinez & Katie Danielson (2016) Emerging reading and the social practice turn in literacy: Still becoming a nation of readers. *Journal of education* 196:3, 19–25. <https://doi.org/10.1177/002205741619600304>
- Hornby, Nick (2006) *The complete polysyllabic spree*. Lontoo: Penguin Books.
- Kupiainen, Reijo (2017) Lukutaidon jälkeen? Teoksessa Vesa Korhonen & Johanna Annala & Pirjo Kulju (toim.) *Kehittämisen palat, yhteisöjen salat*. Tampere: Tampere University Press, 205–218. https://trepo.tuni.fi/bitstream/handle/10024/101964/lukutaidon_jalkeen_2017.pdf
- Kupiainen, Reijo (2021) Monilukutaito muuttaa maailmaa: The New London Group. *Aikuiskasvatus* 41:1, 73–76. <https://doi.org/10.33336/aik.107394>
- Lankshear, Colin & Michele Knobel (2007) Researching new literacies: Web 2.0 practices and insider perspectives. *E-learning* 4:3, 224–240.
- Li, Xiaojing & Roujia Hu (2022) Developing and validating the digital skills scale for school children (DSS-SC) *Information, communication & society* 25:10, 1365–1382. <https://doi.org/10.1080/1369118X.2020.1864002>
- Limberg, Louise & Olof Sundin & Sanna Talja (2012) Three theoretical perspectives on information literacy. *Human IT* 11:2, 93–130. <https://humanit.hb.se/article/view/69>
- Moje, Elizabeth Birr (2015) Doing and teaching disciplinary literacy with adolescent learners: A social and cultural enterprise. *Harvard educational review* 85:2, 254–278. <https://doi.org/10.17763/0017-8055.85.2.254>
- New London Group (1996) A pedagogy of multiliteracies: Designing social futures. *Harvard educational review* 66:1, 60–92. <https://english.utk.edu/wp-content/uploads/2019/08/new-london-group.pdf>
- Mäkinen, Ilkka (1997) *Nödvändighet af LainaKirjasto: Modernin lukuhulun tulo Suomeen ja lukemisen instituutiot*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Mäkinen, Ilkka (2018) Lukuhulun paluu: motivaation käsite suomalaisessa lukemistutkimuksessa toisen maailmansodan jälkeen. *Kasvatus & aika* 12:2, 7–20. <https://journal.fi/kasvatusjaaika/article/view/74127>
- Persson, Magnus (2016) The hidden foundations of critical reading. Teoksessa Paulette M. Rothbauer & Kjell Ivar Skjerdingsstad & Lynne McKechnie & Knut Oterholm (toim.) *Plotting the reading experience: Theory/practice/politics*. Waterloo, ON: Wilfrid Laurier University Press, 19–36.
- Rosen, Matthew (2015) Ethnographies of reading: Beyond literacy and books. *Anthropological quarterly* 88:4, 1059–1084. <https://www.jstor.org/stable/43955502>
- Ross, Catherine Sheldrick (1987) Metaphors of reading. *Journal of library history* 22:2, 147–163. <https://www.jstor.org/stable/25541794>

- Ross, Catherine Sheldrick (2009) Reader on top: Public libraries, pleasure reading, and models of reading. *Library trends* 57:4, 632–656. doi:10.1353/lib.o.0059.
- Ross, Catherine Sheldrick & Lynne McKechnie & Paulette M. Rothbauer (2018) *Reading still matters: What the research reveals about reading, libraries and community*. Santa Barbara: Libraries Unlimited.
- Smagorinsky, Peter (2009) The cultural practice of reading and the standardized assessment of reading instruction: When incommensurate worlds collide. *Educational researcher* 38:7, 522–527. doi:10.3102/0013189X09347583
- Street, Brian (2003) What’s “new” in new literacy studies? Critical approaches to literacy in theory and practice. *Current issues in comparative education* 5:2, 77–91. https://www.tc.columbia.edu/cice/pdf/25734_5_2_street.pdf
- Talja, Sanna (2005) The social and discursive construction of computing skills. *Journal of the American society for information science and technology* 56:1, 13–22. <https://doi.org/10.1002/asi.20091>
- Talja, Sanna (2010) Jean Lave’s practice theory. Teoksessa Gloria J. Leckie & Lisa M. Given & John E. Buschman (toim.) *Critical theory for library and information science: Exploring the social from across the disciplines*. Santa Barbara: Libraries Unlimited, 205–220.
- Talja, Sanna & Annemaree Lloyd (2010) Challenges for future research on learning, literacies and information practices. Teoksessa Annemaree Lloyd & Sanna Talja (toim.) *Practicing information literacy: Bringing theories of learning, practice and information literacy together*. Lontoo: Chandos, 357–364.
- Talja, Sanna & Pertti Vakkari & Jenny Fry & Paul Wouters (2007) The impact of research cultures on the use of digital library resources. *Journal of the American society for information science and technology* 58:11, 1674–1685. <https://doi.org/10.1002/asi.20650>
- Tattersall Wallin Elisa & Jan Nolin (2020) Time to read: Exploring the timespaces of subscription-based audiobooks. *New media & society* 22:3, 470–488. doi:10.1177/1461444819864691
- Trasmundi, Sarah Bro & Lydia Kokkola & Theresa Schilhab & Anne Mangen (2021) A distributed perspective on reading: implications for education. *Language sciences* 84:101367. <https://doi.org/10.1016/j.langsci.2021.101367>
- Tuominen, Kimmo & Reijo Savolainen & Sanna Talja (2005) Information literacy as a sociotechnical practice. *The library quarterly* 75:3, 329–345. <https://doi.org/10.1086/497311>
- Vakkari, Pertti (1983) *Kirjojen lukeminen ja yleiset kirjastot*. Helsinki: Kirjastopalvelu.
- Wolf, Maryanne (2019) Silmäilystä on tullut normaali tapa lukea. *Niin & näin* 2019:1, 35–38. <https://netn.fi/sites/www.netn.fi/files/netn191-07.pdf>

2.3 Tehtävälähtöinen informaatiovuorovaikutus

Sanna Kumpulainen ja Mika Sihvonon

- Informaatiota ei käytetä pelkästään informaation vuoksi, vaan yleensä informaatio liittyy johonkin muuhun ihmisten toimintoon tai tehtävään.
- Informaation ja informaatiojärjestelmien hyödyllisyyttä arvioidaan sen mukaan, kuinka hyvin ne auttavat tehtävän tavoitteiden saavuttamisessa.
- Tiedonhakua, tiedon valintaa ja tiedonkäyttöä jäsennetään tehtävän avulla.
- Tehtävälähtöisessä informaatiovuorovaikutuksessa voidaan eritellä toimintoja eli aktiviteetteja, joissa vuorovaikutus informaation kanssa on tietyn tyyppistä.
- Malli auttaa jäsentämään monimutkaista kokonaisuutta.

Tiedonhaun tutkimuksessa ollaan kiinnostuneita siitä, miten erilaiset tietojärjestelmät tuottavat relevantteja dokumentteja. Tiedonhaku on informaatiotutkimuksen keskeisiä tutkimusalueita ja liittyy ilmiönä kaikkiin nykyisiin digitaalisiin informaatioympäristöihin. Tiedon äärellä hakeutuminen sisältää nykyisin hyvin usein tiedonhakuja erilaisista digitaalisista järjestelmistä. Tiedonhaun tutkimuksessa näkökulma on kuitenkin suppeahko.

Tehtävälähtöinen informaatiovuorovaikutus on puolestaan kiinnostunut siitä, miten eri tavoin ihmiset käyttävät erilaisia tiedonlähteitä, tietoresursseja ja tiedonhakuja järjestelmiä jonkin inhimillisen tehtävän tavoitteiden saavuttamiseksi. Tehtävä tarkoittaa ihmisen päämäärätietoista toimintaa. Tehtävää tehdessä syntyy tiedontarpeita, joita pyritään mahdollisuuksien mukaan täyttämään etsimällä

tietoa erilaisista lähteistä, ja tehtävä on siten keskeinen tiedonhankinnan ja tiedonhaun toimintojen käynnistäjä.

Informaatiotutkimus tieteenalana on kiinnostunut dokumentteihin tallennetusta tiedosta ja sen järjestämisestä, tiedon kanavien äärelle hakeutumisesta, dokumentteihin tallennetun tiedon etsinnästä eri muodoissaan sekä tiedon hyödyntämisen ilmiöistä. Näitä ilmiöitä on lähestytty makro- ja mikrotason näkökulmista. Makrotason tutkimus on luonteeltaan yhteiskuntatieteellistä, ja siinä ollaan kiinnostuneita ihmisryhmien toiminnasta. Mikrotasolla on tarkasteltu perinteisesti tiedonhakujärjestelmän toimintaa tai toiminnan osaa, esimerkiksi tiettyä kokoelmaa. Ensimmäistä, yhteiskuntatieteellistä lähestymistapaa kutsutaan tiedonhankintatutkimukseksi. Jälkimmäistä, järjestelmiin pureutuvaa näkökulmaa kutsutaan tiedonhaun tutkimukseksi. Tehtävälähtöisessä informaatiovuorovaikutuksessa nämä kaksi aiemmin erillistä, mutta samoja ilmiökenttiä tarkastelevaa tieteenalaa on tuotu yhteen.¹ Nykyään tiedonhankintatutkimus on lähentynyt tiedonhaun tutkimusta informaatioympäristöjen digitalisaation mukana.

Ihmisen ja informaation välinen vuorovaikutus (*human information interaction*) syntyi terminä vastapariksi ihmisen ja tietokoneen väliselle vuorovaikutukselle (*human computer interaction*).² Informaatiovuorovaikutuksella tarkoitetaan nimenomaan dokumentteihin tallennetun tietosisällön kanssa tapahtuvaa vuorovaikutusta. Lisäksi termejä tiedonhakuvuorovaikutus (*information retrieval interaction*)³ ja vuorovaikutteinen tiedonhaku (*interactive information retrieval*)⁴ käytetään jokseenkin samasta ilmiökentästä, tosin näissä usein ilmiö nähdään rajatumpana, usein kyselyiden ja hakukoneen vuoropuheluna (lähikäsitteistä lisää, ks. luku 1.2).

1 Ingwersen & Järvelin 2005.

2 Blandford & Attfield 2010.

3 Ingwersen 1992.

4 Ruthven 2009.

HCI ja HII

Ihmisen ja tietokoneen välinen vuorovaikutus (*Human-Computer Interaction*, HCI) on tutkimusalue, joka tarkastelee ihmisen ja tietokoneen vuorovaikutusta eri näkökulmista. Tietokoneen graafisen käyttöliittymän suunnitteleminen käyttäjää miellyttävällä tavalla korostui erityisesti kotitietokoneiden markkinoinnin alkuaikoina, jolloin ihmisen ja tietokoneen vuorovaikutuksen keskeiset kysymykset jäivät taka-alalle. HCI-tutkimuksessa pyritäänkin huomioimaan ihmisen ja tietokoneperustaisen laitteen vuorovaikutuksen monet ulottuvuudet, kuten psykologiset tekijät, työskentelykäytännöt ja terveysvaarat.⁵

Ihmisen ja informaation välisen vuorovaikutuksen (*Human Information Interaction*, HII) tutkimus ulottuu monitieteellisesti muun muassa niihin lähestymistapoihin, jotka tarkastelevat informaatioon liittyviä sosiaalisia, hallinnollisia ja teknologisia kysymyksiä. Digitaalisen informaation yleistyessä HII liittyy enenevästi myös HCI:n tutkimusalueisiin, kun ihmisen on kyettävä tulkitsemaan tietokoneen esittämä informaatio. HII-alueen tutkimusta voidaan tehdä puhtaasti kognitiivisella tasolla, ottamatta kantaa informaatiota välittävään teknologiaan. Keskeistä HII-tutkimuksessa on ihmiselle välittyvän informaation osuus päätöksenteossa.⁶ Nykyisin tiedonhaun tutkimus on lähentynyt datatiedettä ja tekoälytutkimusta, ja perinteinen informaatiotutkimuksen puolella tehty tekstitiedonhaun tutkimus on jäänyt vähemmälle. Datatiede ja tekoälytutkimus lähtevät liikkeelle dokumenteista ja kokoelmien ominaisuuksista, joiden tarkastelukulmana voi olla myös erilaisten informaatioresurssien käyttämisen perusteella syntyvien digitaalisten jälkien, kuten klikkausten ja kyselyiden, mallintaminen. Vastapainona tälle teknologiselle

⁵ Preece ym. 1994, 7.

⁶ Fidel 2012, 18–20.

intressille informaatiotutkimuksessa tehtyä tiedonhaun tutkimusta leimaa pyrkimys ymmärtää ihmisen toimintaa nykyajan monimutkaisessa ja digitaalisessa informaatioympäristössä. Tutkimuksen taustalla on myös tarve tukea ihmisten toimintaa eri tavoin, esimerkiksi järjestämällä tietoresursseja sekä kehittämällä uusia tiedonhakutapoja ja tiedonhakujärjestelmiä.

Tehtävä ja tiedontarve

Tiedonhaun ja tiedonhankinnan tavoitteena on tiedontarpeiden tyydyttäminen.⁷ Tiedontarpeita on käsitteellistetty monin eri tavoin. Se, mistä tiedontarpeet kumpuavat, on tiedonhaun tutkimuksissa usein epämääräinen ja jäsentymätön kenttä, eräänlainen musta laatikko. Tiedonhankintatutkimuksessa tätä mustaa laatikkoa selitetään laajasti, mutta usein se jää irralleen varsinaisesta tiedonhausta ja sen prosesseista. Tehtävälähtöisessä lähestymistavassa tiedonhaun ja hankinnan ilmiöihin liitettävät tiedontarpeet selitetään laajemman, tavoitteellisen toiminnan päämäärien kautta.⁸ Tiedonhankinta ja -haku, tiedon käyttö ja hyödyntäminen sekä edelleen jakaminen ovat sidotut laajemman toiminnan kontekstiin eli esimerkiksi osaksi tehtävän suorittamisen prosessia ja osaksi erilaisia organisatonaalisia ja kulttuurisia ilmiöitä.

Ihmiset hakevat tietoa, jotta suoriutuisivat paremmin käsillä olevista tehtävistä.⁹ Tehtävä voi olla luonteeltaan millainen tahansa, esimerkiksi lomamatkan suunnittelu tai yrityksen tulevan vuosisuunnitelman tekeminen. Tehtävien voidaan sanoa virittävän tiedontarpeen tai tietämyksen eräänlaiseen puutostilaan (ASK, *anomalous state of knowledge*).¹⁰ ASK kuvaa tilannetta, jossa jokin henkilön oman tietämyksen rakenteen ristiriita tai epävarmuus aiheuttaa

7 Taylor 1968.

8 Byström & Kumpulainen 2020.

9 Kumpulainen 2013, 16.

10 Belkin ym. 1982.

tiedon puutostilan. Tiedontarve tai epämääräisemmin aistittu tiedon puutostila puolestaan synnyttää toimintaa eli tiedonhankintaa, jonka osana muun muassa tutulta ihmiseltä kysymisen lisäksi tapahtuu myös tiedonhakua.¹¹ Tiedontarve on vaikeasti hahmotettava ja tutkittava käsite, mutta se voidaan johtaa sosiokulttuurisesta toimintaympäristöstä ja käsillä olevan tehtävän suorittamisen kontekstista käytännölliseen tehtävän jäsentämiseen sekä tiedonhankintaan ja -hakuihin.¹² Tiedonhaut tapahtuvat sarjana, eivätkä ne ole yksittäisiä kyselyitä vaan dynaamisia, ihmisen toiminnan ja osaamisen kanssa kehittyviä episodeja. Tämä kompleksinen asetelma olisi hyvä ottaa huomioon, jotta kaikki ilmiöön liittyvä olennainen saadaan kerättyä ja jotta vältetään liian kapeat näkökulmat.

Tehtävälähtöisen tiedonhaun teoria ja prosessimallit

Informaatiotutkimuksen klassikkotutkimus tiedonhaun prosessimallista (*information search process*, ISP) sitoo opiskelijan informaatiotoiminnot tutkielman kirjoittamisen prosessin eri vaiheisiin.¹³ Tästä johdettu uudempi malli on Tampereen yliopistossa kehitetty tehtävälähtöisen tiedonhaun teoria, jossa tehtäväprosessissa nähdään kolme osiota: 1) ennen fokuksen muodostamista, 2) fokuksen muodostamisen aikana ja 3) sen jälkeen.¹⁴ Eri vaiheissa tiedonhaku on varsin erilaista: taustoittavaa ja yleistä, tarkentuvaa ja viimeisessä vaiheessa yksityiskohtaista faktojen hakemista.¹⁵ Tätä voidaan selittää sillä, että jos ymmärrys taustoittavasta toiminnasta on jäsentymätön, on vaikea tuottaa sopivia kyselyitä järjestelmille ja järjestelmän tuotosten arviointi on vaikeaa.¹⁶

Tehtävälähtöisessä tarkastelussa otetaan usein huomioon niin kutsuttu kognitiivinen ulottuvuus, joka käsittää ihmismielen tie-

11 Wilson 1981.

12 Byström & Kumpulainen 2020.

13 Kuhlthau 1991; 1993.

14 Vakkari 2001.

15 Vakkari 2001.

16 Vakkari & Huuskonen 2012.

dollisia toimintoja, kuten muistamisen ja oppimisen, mutta esimerkiksi tunneulottuvuutta ei välttämättä tarkastella lainkaan. Poikkeuksen tästä tekee aiemmin mainittu Kuhlthau'n tiedonhaun malli,¹⁷ jossa tarkastellaan myös affektioita eli tunteita tehtävän suorituksen eri vaiheissa.

Kehittyviä prosesseja tutkittaessa täytyy tarkastella ajallisesti pitkittäisiä tapahtumajaksoja. Tätä voi lähestyä vaihteittain¹⁸ tai ottaamalla laajemman tehtävän tarkastelun alle ja pilkkomalla sen osatehtäviin, joita tarkastellaan lähemmin toimintojen ketjuna.¹⁹

Tehtävän kompleksisuus ja vaikeus

Tavoitteellisen toiminnan eli tehtävän luonnetta voidaan myös karvoittaa suhteessa tiedonhakuun ja tiedonhankintaan. Esimerkiksi tehtävän kompleksisuus kertoo tehtävän monimutkaisuudesta eli siitä, kuinka monimutkaiselta se suorittajasta tuntuu, kun tarkastellaan sitä, kuinka paljon tiedetään tehtävän lopputuloksesta, suorittamistavasta sekä sisällöstä.²⁰ Tehtävän kompleksisuuden on osoitettu vaikuttavan tiedonhankintaan ja -hakuun eri tavoilla ja eri sovellusaloilla.²¹ Toinen samankaltainen termi on tehtävän vaikeus, jossa tarkastellaan ihmisten itse raportoimaa koettua vaikeusastetta tai hankaluutta.²² Joskus kokeellisissa vuorovaikutteisen tiedonhaun tutkimusasetelmissä, joissa käytetään etukäteen muotoiltuja haku-tehtäviä, on käytetty monimutkaisuuden mallintamisessa kognitiivisen kompleksisuuden asteikkoa.²³ Tämä perustuu opetuksessa ja osaamistehtävien arvioinnissa paljon käytettyyn taksonomiaan.²⁴

17 Kuhlthau 1991; 1993.

18 Ks. esim. Kuhlthau 1993; Vakkari 2001.

19 Ks. esim. Järvelin ym. 2015; Kumpulainen, 2014; Lin & Belkin 2000.

20 Byström & Järvelin 1995.

21 Esim. Byström & Järvelin 1995; Kumpulainen & Järvelin 2010; Saastamoinen ym. 2013; Saastamoinen & Järvelin 2017.

22 Wildemuth ym. 2014.

23 Arguello ym. 2012; Kelly ym. 2015.

24 Ks. esim. Krathwohl 2002.

Siinä kukin ajattelun taso vastaa jotain toimintaa. Alin taso käsittää toiminnon muistaa, toinen taso toiminnon ymmärtää, kolmas taso toiminnon soveltaa, neljäs taso toiminnon analysoida. Ylimmät, haastavimmat tasot, viides ja kuudes, sisältävät toiminnot arvioida ja luoda.

Tutkimusmenetelmistä

Koska tehtävälähtöisessä informaatiovuorovaikutuksen tutkimuksessa ollaan kiinnostuneita juuri ihmisten tavoitteellisesta toiminnasta, täytyy se ottaa myös tutkimusasetelmissä huomioon.²⁵ Usein tehtävälähtöisissä tutkimusasetelmissä kerätään tutkimusaineistoja monin eri keinoin, jolloin puhutaan menetelmä- tai aineistotriangulaatiosta. Se parantaa monesti analyysin tarkkuutta ja luotettavuutta. Jotta ihmisen toiminta saadaan kontekstissaan kuvattua, menetelmät sisältävät usein haastatteluita, havainnointeja ja muita etnografisia menetelmiä, kuten kyselyitä ja päiväkirjoja sekä näiden yhdistelmiä. Myös järjestelmien lokiaineistoja on käytetty yhdistettyinä näihin menetelmiin.²⁶

Haastattelu

Fujiin²⁷ mukaan haastattelut ovat paras tapa saada syvällistä tietoa siitä, kuinka ihmiset havaitsevat ja kokevat erilaiset ilmiöt päivittäisessä elämässään. Haastatteluilla saadaan vastauksia esimerkiksi miksi-kysymyksiin, joilla voidaan selittää jotain toimintamuotoa tai -tapaa, kuten tiedonhakupajoja. Toki haastatteluissa saadaan haastateltavan näkökulmasta tavallaan kertomuksen läpi suodatettu versio, eivätkä ihmiset aina ole halukkaita tai kykeneväisiä paljastamaan aivan kaikkea, varsinkaan jos se asettaa heidän osaamisensa tai tietonsa kyseenalaiseen valoon. Haastattelut ovat kuitenkin erittäin hyvä ja kustannustehokas tutkimusmenetelmä, jolla saadaan rikasta

25 Ks. esim. Kumpulainen 2017.

26 Esim. Kumpulainen & Järvelin 2010.

27 Fujii 2017.

ja vivahteikasta tutkimusaineistoa. Haastattelututkimuksen aineiston analyysi tapahtuu usein laadullisia, tyypitteleviä menetelmiä käyttäen.

Tyypillisimmillään tehtävälähtöisissä informaatiovuorovaikutuksen haastattelututkimuksissa käytetään puolistrukturoitua haastattelua, jossa haastattelurunko antaa joustavan rakenteen haastattelun etenemiselle. Yksi esimerkki haastattelututkimuksesta on Tampereen yliopistossa toteutettu historian tutkijoiden tehtävälähtöisen tiedonhaun tutkimus, jossa tarkasteltiin historiallisten sanomalehtiaineistojen käyttöä.²⁸

Havainnointi

Havainnointi on hyvä tapa saada kerättyä tietoa myös tiedostamattomista toiminnoista, joista tehtävän suorittajat eivät välttämättä ole tietoisia tai joita he eivät pysty sanoittamaan. Yksi havainnoinnin tapa on varjostaminen, jossa havainnoija voi kysellä ja keskustella vapaasti, muttei osallistu suoraan tehtävän suorittamiseen vaan tarkkailee toimintaa.²⁹ Havainnointia on käytetty muun muassa sosiaaliryöön kontekstissa³⁰ ja monialaisessa tutkimuksessa³¹ sekä bioteknologien informaatiovuorovaikutuksen esteiden tutkimuksessa.³²

Kysely

Kysely on taloudellinen tiedonkeräystapa saada tietoja ilmiöistä, joista jo jonkin verran tiedetään. Jos ilmiöstä ei tiedetä mitään, voi kyselyn suunnittelu olla haastavaa ja vastausten kontrollointi vaikeaa. Tästä huolimatta kysely sopii myös tutkimusalueille, joissa luottamuksellisuus ja anonymiteetti ovat keskeisiä: se on käytännössä ainoa tapa kerätä tietoa, jossa mitään jälkiä vastaajista ei jää, ellei niin haluta. Kyselyissä voidaan kerätä avoimia vastauksia tai ky-

28 Late & Kumpulainen 2021.

29 Kumpulainen ym. 2009; McDonald 2005.

30 Huuskonen & Vakkari 2013.

31 Saastamoinen & Järvelin 2017.

32 Kumpulainen & Järvelin 2012.

sely voi olla rakenteisempi. Avointen vastausten analyysi on hyvin pitkälle laadullista, kun taas rakenteisemmissa kyselyissä vastausvaihtoehtoja annetaan valmiina. Esimerkkinä tällaisesta voidaan mainita Likert-asteikollinen kysely, jossa kysytään esimerkiksi kokemuksia ja mielipiteitä asioihin skaalalla täysin samaa mieltä – täysin eri mieltä, ja vastaukset koodataan analyysissä numeeriselle järjestyksasteikolle viidestä yhteen.

Kyselyillä voidaan kerätä myös muun muassa erilaisia frekvenssitilastoja. Tällöin analyysissä käytetään tilastollisia ja kvantitatiivisia menetelmiä. Kyselytutkimuksia käytetään esimerkiksi tietyn palvelun tai tietoresurssin käyttäjäkyselyissä. Kyselyä voidaan käyttää myös kartoittavana menetelmänä, jossa haarukoidaan joitain ilmiöitä, kuten jonkin ihmisryhmän informaatioympäristöä. Kyselyä on hyödynnetty muun muassa Tampereen yliopistossa molekyyliiläketieteilijöiden informaatioympäristöjä tutkiessa.³³

Päiväkirja

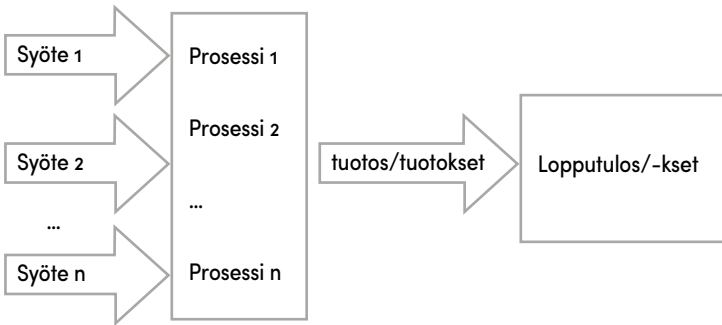
Päiväkirjatutkimuksissa osallistujia pyydetään täyttämään joko digitaalisia tai paperisia päiväkirjoja omasta informaatiokäyttäytymisestään ja -vuorovaikutuksestaan sekä näihin liittyvistä kokemuksistaan. Päiväkirja voi olla valmiiksi rakenteinen, eli siinä esitetään kysymyksiä, kuten ”Mitä tietoa hait?” Se voi olla myös vapaamuotoisempi kirjoitelma aiheesta, jolloin lähestytään kertomusmuotoista esitystapaa. Päiväkirja voi kattaa esimerkiksi yhden tilanteen tai yhden päivän tai jopa hyvinkin pitkän ajanjakson. Päiväkirjojen haasteena on se, että osallistujat eivät välttämättä aina muista merkitä kokemuksiaan ylös, mistä on hyvä muistuttaa osallistujia. Päiväkirjoja on käytetty muun muassa tutkittaessa kaupunkien hallinnon työtä³⁴, tohtoriopiskelijoiden tiedonhaun serendipisyyttä eli yllättäviä sattumia³⁵ ja opiskelijoiden yhteistyötä tiedonhaussa.³⁶

33 Roos ym. 2008.

34 Byström & Järvelin 1995; Saastamoinen & Kumpulainen 2014.

35 Sun ym. 2011.

36 Shah & Leeder 2015.



Kuvio 1. Ohjelmateorian yleinen esitys.

Tehtävälähtöisen informaatiovuorovaikutuksen evaluointimalli

Tarkasteltaessa sitä, kuinka hyvin erilaiset tietoresurssit ja tiedonhakujärjestelmät tukevat ihmisen toimintaa, puhutaan evaluoinnista. Evaluointia voidaan tehdä tiedonhakujärjestelmien eri osien näkökulmasta, mutta myös laajemmin ihmisen informaatiotoimintojen näkökulmasta. Yksi malleista, joka pyrkii jäsentämään tätä ilmiökenttää, on tehtävälähtöisen informaatiovuorovaikutuksen evaluointimalli.³⁷ Mallissa tarkastellaan eri informaatiotoimintoja eli -aktiviteetteja oppimistehtävän näkökulmasta, mutta malli sopii myös muiden tehtävien tarkasteluun yleisen luonteensa vuoksi.

Tehtävälähtöisen informaatiovuorovaikutuksen (*task-based information interaction*, TBII) mallin aktiviteetit ovat:

- tehtävän suunnittelu ja reflektiivinen arviointi
- informaatioyksiköiden hakeminen
- informaatioyksiköiden valinta
- informaatioyksiköiden kanssa työskentely
- syntetisointi ja raportointi

37 Järvelin ym. 2015.

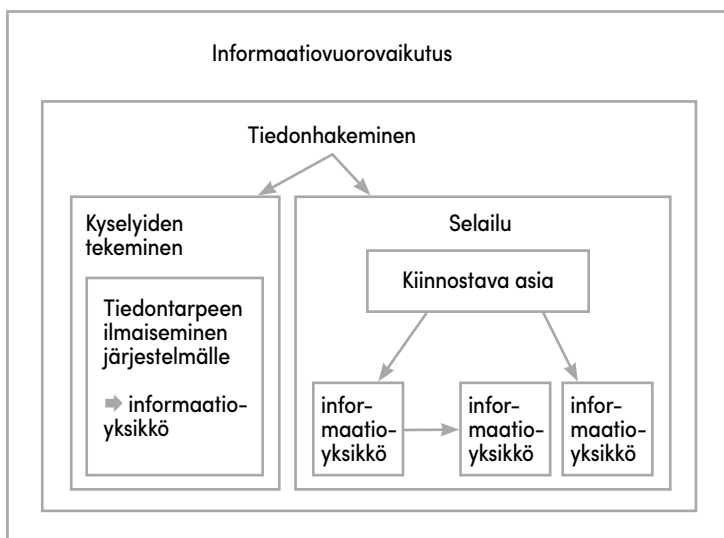
Informaatiovuorovaikutus käsitetään huomattavasti laajempänä ilmiönä kuin pelkkä tiedonhaku. Se, miten informaatioyksiköiden kanssa työskennellään, vaikuttaa niiden hyödylliseksi kokemiseen. Informaatioyksiköillä tarkoitetaan jossain muodossa tallennetun tiedon merkityksellistä osasta, joka voi olla periaatteessa koko dokumentti, kirja, kuva tai piirros, äänite tai sen pätkä tai muu informaatiokappale. Malli pohjautuu yhteiskunnallisten kehittämis- ja toimenpideohjelmien evaluointiin, jossa pyritään selittämään, toimiiko jokin toimeenpantu ohjelma ja miten se toimii ja siten selittämään kausaalisia kuvauksia ohjelman vaikuttavuudesta.³⁸ Yleinen esitys ohjelmateorialle löytyy kuviosta 1. Mallissa kullekin aktiviteetille on kuvattu syötteet, prosessi, tuotokset ja lopputulokset. TBII-malli esittää edellä mainittujen informaatioaktiviteettien evaluointin ohjelmateorian kautta ja kokoaa lopuksi näiden yksittäisten aktiviteettien evaluointikehikot yhteen yleiseksi evaluointikehikoksi koko tehtävälähtöisen informaatiovuorovaikutuksen prosessille.

Seuraavaksi aktiviteetit esitetään yleisellä tasolla ja tämän jälkeen esitellään koko malli, jossa nämä yksittäiset aktiviteetit kootaan yhteen. Malli on yleinen ja kattaa laajasti erityyppisiä informaatioaktiviteetteja, minkä ansiosta sen avulla voidaan selittää monipuolisesti eri informaatioaktiviteettien välisiä eroja ja yhteen kietoutumista. Mallin hyödyllisyys korostuu, kun tarkastellaan erilaisissa todellisissa toiminta- ja informaatioympäristöissä tapahtuvia erilaisia informaatioaktiviteetteja.

Tehtävän suunnittelu

Tehtävälähtöiselle informaatiovuorovaikutukselle on keskeistä tarkastella nimenomaan tehtävän ja sen ominaisuuksien vaikutusta tiedonhakuun, tietoresurssien valintaan sekä niiden kanssa työskentelyyn ja niin edelleen. Tehtävä tarkoittaa siis jotain tavoitteellista tehtävää, kuten opiskelu- tai työtehtävää. TBII-mallissa tehtävän suunnittelu ja reflektiivinen arviointi nähdään meta-

38 Rossi ym. 2003.



Kuvio 2. Kyselyiden tekeminen ja selailu informaatioyksiköitä haettaessa.

kognitiiviseksi muita aktiviteetteja tukevaksi toiminnoksi. Tähän aktiviteettiin viitataan jatkossa lyhennetysti tehtävän suunnitteluna. Tehtävän suunnittelun aikana aktiviteetit järjestetään suunnitelmaksi tehtävän tavoitteen saavuttamiseksi ja yksittäisten aktiviteettien sisällöt päätetään kulloisenkin tehtävästä muodostetun ymmärryksen mukaan. Tämä aktiviteetti on luonteeltaan dynaaminen, ja se sopeutuu ja kehittyy tehtävän suorituksen aikana.³⁹ Jos tehtävänä olisi vaikkapa lomamatkan suunnittelu, saattaisi tiedontarpeena olla hinnan, ajankohdan, harrastusmahdollisuuksien ja kohdepaikan ilmaston selvittäminen. Tehtävän aikana ymmärrys tehtävän olemuksesta sekä siinä tarvittavasta tiedosta kehittyy ja tehtävän tekijä eli toimija kohtaa suorituksen aikana erikoistarjouksia, uusia kohteita, kavereiden ja järjestelmien suosituksia sekä mo-

39 Vrt. fokuksen muodostaminen, Kuhlthau 1993; Vakkari 2001.

nenlaisia muita informaatioyksiköitä, joiden perusteella hän tekee uusia valintoja tavoitteen saavuttamiseksi.

Tehtävän suunnittelussa prosessiin otetaan syötteitä. Syötteitä ovat 1) erilaiset kontekstuaaliset seikat, kuten aika, työympäristö ja sosio-organisaationaaliset tekijät, 2) havaittu tehtävä sellaisena kuin tekijä sen sillä hetkellä hahmottaa sekä 3) informaatioympäristö. Prosesseina aktiviteetissa ovat tehtävän suunnittelu, monitorointi ja arviointi. Nämä prosessit kohdistuvat sekä tehtävään että sen suorittamisen tapaan. Prosessin tuotoksina ovat ymmärrys tehtävästä eli suunnitelma, tehtävän tavoitteet ja valitut aktiviteetit eli etenemistavat. Myös kannattavuuslaskelma on yksi prosessin tuotos. Siinä punnitaan, kuinka suuri vaivannäkö ja ponnistelu on tarkoituksenmukaista halutun asiantilan saavuttamiseksi. Tehtävän suunnittelu ja reflektiivinen aktiviteetti ovat toistuvia ja vaikuttavat kaikkiin informaatiovuorovaikutuksen vaiheisiin ja osiin. Tehtävän suunnittelussa on mielenkiintoista erityisesti sen vaikutus koko tehtävän suorittamiseen ja yksittäisiin muihin aktiviteetteihin. Tämän perusteella on mahdollista tutkia, miten erilaisten tehtävästrategioiden valinta vaikuttaa eri aktiviteettiosasten tuotosten tyyppiin ja laatuun.

Tietoyksiköiden hakeminen

Tietoyksiköiden hakemisen aikana toimija on vuorovaikutuksessa tiedonhakujärjestelmän kanssa tarkoituksenaan saavuttaa tehtävän suorittamisen kannalta merkityksellinen tavoite. Tietoyksiköiden hakeminen on välineellinen toiminto, jolla pyritään edistämään tehtävää ja sen suorittamista. Kun informaatiovuorovaikutus on kokonaisuudessaan kestoaltaan varsin pitkä prosessi ja saattaa kestää viikkoja tai jopa kuukausia, riippuen toki tehtävästä, tiedon hakeminen on kestoaltaan huomattavasti lyhyempi aktiviteetti ja toistuu useita kertoja eri kohdissa informaatiovuorovaikutusprosessia. Tietoyksiköiden hakeminen voidaan jakaa kyselyiden tekoon ja selailuun (ks. kuvio 2). Kyselyn tekeminen koostuu tiedontarpeen ilmaisemisesta järjestelmälle, kun taas selailu sisältää sarjan, jossa ensin nähdään jostain kiinnostavaa esimerkiksi silmäillessä tietoyksiköitä tai tuloslis-

taa, minkä jälkeen huomio varsinaisesti kiinnitetään informaatioyksiköihin. Lopuksi tietosisällöt otetaan haltuun tai hylätään niiden tarkastelun jälkeen. Tietoyksiköiden hakemisen tuotokset toimivat syötteinä muille aktiviteeteille, samaan tapaan kuin tehtävän suunnittelussa.

Kun päätös tiedon hakemisesta on syntynyt tehtävän suunnittelun yhteydessä, täytyy tehtävä –sellaisena kuin se sillä hetkellä ymmärretään – muuttaa hakuilmauksiksi. Tämä johtaa tietoyksiköiden tarkasteluun. Tiedontarpeiden ymmärrys muuttuu ja kehittyy tiedonhaun aikana tietojärjestelmistä saadun tiedon vaikutuksesta. Syötteenä prosessissa toimivat kontekstuaaliset tekijät (ks. edellinen alaluku), havaitut tehtävytarpeet sekä tiedonhakijan eli toimijan ominaisuudet. Prosessiin kuuluvat tiedonhaun kanavien valinta, tiedonhaun tarpeen ilmaiseminen ja muotoilu, huomion kiinnittäminen informaatioyksiköihin sekä teknologisen ympäristön prosessit. Tuotoksena on hakutulostilaus, joka palvelee tehtävässä tarvittavan informaation saamista.

Informaatioyksiköiden valinta

Tietoyksiköiden hakemisen jälkeen täytyy valita tehtävään sopivat informaatioyksiköt. Valintapäätöstä tehtäessä on arvioitava informaation hyödyllisyys tai relevanssi. Relevanssin käsitteestä kerrotaan tarkemmin luvussa 3.2. Tämä aktiviteetti on usein kietoutunut tiedonhakuun, ja käytännössä sitä voi olla vaikea erottaa tiedon hakemisesta omaksi aktiviteetiksi.⁴⁰ Ilmiön analyttinen tarkastelu on kuitenkin mahdollista, ja se avaa aktiviteetin tiettyjen ominaisuuksien tarkastelulle oman käsitteellisen kehyksen.

Valinnan edellytyksenä on vuorovaikutus informaatioyksiköstä tuotetun tai tarjotun esityksen kanssa. Näitä voivat olla esimerkiksi otsikot ja abstraktit, otteet ja pätkät tekstistä, metadata ja avainsanat tai kokotekstiesitykset. Tämän vuorovaikutuksen tarkoituksena on arvioida, mitkä informaatioyksiköt sopivat tehtävän suoritukseen

40 Korkeamäki & Kumpulainen 2019.

ja kattavat riittävän osan tehtävän vaatimasta tietomäärästä. Tässä yhteydessä puhutaan relevanssin arvioinnista. Loput tarjolle saadut informaatioyksiköt jätetään huomiotta. Erilaiset toimijat arvioivat relevanssia varsin eri tavoin riippuen esimerkiksi aiemmasta aihe-tietämyksestä, taidoista tai oppimistyyleistä.

Informaatiovuorovaikutuksen arviointikehikossa valinta-akti-viteetti saa syötteenä kontekstitekijöitä, kulloisenkin ymmärryksen tehtävästä ja sen vaiheesta, toimijan henkilökohtaiset ominaisuudet sekä edellisestä aktiviteetista tuotoksina syntyneet hakutulokset jos-sain järjestelmän tarjoamassa järjestyksessä ja esitysmuodossa. Pro-ssessi sisältää kolme toimintoa: 1) informaatioyksiköiden esitysten valinta, 2) näiden tarkastelu ja muokkaus sekä 3) päätöksen teko. Tuotoksina syntyy joukko, joka sisältää valitut informaatioyksiköt, jotka ovat sopivia tehtävän suoritukseen. Loput informaatioyksiköt hylätään. Lopputuloksena on informaatiota, joka auttaa tehtävän suorituksessa, sekä ylipäätään parantunut käsitys saatavilla olevasta informaatiosta.

Informaatioyksiköiden kanssa työskentely

Tehtävän suoritus edellyttää usein informaatioyksiköiden hankintaa ja tulkintaa. Tiedonhankintaa tehdään nykyaikana pääosin interne-tissä, jossa informaatio on välittömästi saatavilla ja sen haltuunotto alkaa jo samanaikaisesti hakutuloslistauksen ja linkkitietojen luke-misen kanssa. Informaatioyksiköitä luetaan ja tulkitaan ikään kuin tehtävän linssien läpi. Informaatiota valitaan, kerätään, organisoi-daan, annotoidaan, mikä tarkoittaa, että siihen lisätään kuvaavia tunnisteita, ja lisäksi informaatiota yhdistellään myöhempää käyt-töä varten.

Sähköisen dokumentin selaamisessa kaikki selailustrategiat joh-tavat lopulta jonkin semanttisen komponentin lukemiseen, joka voi olla vaikka pätkä tekstiä tai kuva.⁴¹ Jokaisen tiedonlähteen ja informaatioyksikön informaatioarkkitehtuuri ohjaa informaatio-

41 Arvola & Kekäläinen 2010.

yksiköiden kanssa työskentelyä. Tarjolla voi olla esimerkiksi erilaisia semanttisia jäsennyksiä sekä erilaisia tekstin rakenteita ja osasia (esimerkiksi otsikot, asiasanoitus tai fasettihaku). Informaatioyksiköiden yhteydessä saattaa olla visuaalisia tai sanallisia vihjeitä, jotka ohjaavat informaatioasisältöjen äärelle. Informaatioyksiköiden kanssa työskentelyn aikana toimija saattaa esimerkiksi kopioida tietosisältöjä tai tehdä muistiinpanoja tai merkintöjä dokumentteihin.

Informaatiota tulkitaan tiedon hyödyntäjän eli toimijan tehtävästä luoman ymmärryksen puitteissa.⁴² Informaatioasisältöjen läpikäyminen saattaa osoittaa uusia hakuavaimia, joita voidaan käyttää uusien tiedonhakujen tekemiseen. Tällöin palataan takaisin tiedon hakemisen aktiviteettiin ja jatketaan prosessia siitä. Itse kukin on varmasti tietoa hakiessaan löytänyt uusia, asiaan liittyviä informaatioasisältöjä, joita voidaan hyödyntää uusissa hauissa. Esimerkiksi aiemmassa kappaleessa kuvatussa lomamatkan suunnittelutehtävässä uudet hakuavaimet voivat olla aiempien hakujen yhteydessä ilmi tulleita tietoja matkakohteiden arviointisivustoista tai parempia tuloksia tuottavia hakufraaseja.

Informaatioyksiköiden kanssa työskentely ottaa syötteenä vastaan kontekstitekijät sekä toimijan piirteet, kuten asiantuntemuksen tason ja havaitut tehtävän tarpeet. Myös edellisen prosessin osatekijät, kuten informaatioyksiköiden valinta ja saadut tuotokset eli valitut informaatioyksiköt, ovat syötteitä. Prosessissa luetaan, organisoidaan ja jäsennetään dokumentteja. Niihin tehdään muistiinpanoja ja annotaatioita eli alkuperäiseen dokumenttiin tai sen yhteyteen tuotetaan omia merkintöjä. Tuotoksena ovat tunnistetut semanttiset eli tehtävän kannalta merkitykselliset osat, jotka lisäävät ymmärrystä informaatioyksiköistä ja siten tuottavat lopputuloksena kohentuneen ymmärryksen tehtävästä.

Kiinnostavia tutkimuskohteita tässä aktiviteetissa ovat muun muassa lukutavat (esimerkiksi lineaarinen vs. selektiivinen) ja erilaiset informaation esitystavat (esimerkiksi laite, selain ja sisällön

42 Ingwersen & Järvelin 2005, 274–276.

organisointi). Erityisen kiinnostavaa on erilaisten tietoresurssien ja informaatiotyyppien yhteiskäyttö.⁴³ Kun tarkastelukulma laajenee koskemaan koko informaatioympäristöä, yksittäisen tiedonlähteen tai dokumenttityypin sijaan lähestytään tutkimuksellisesti aitoja tilanteita, mutta tutkimuksen teon haasteet samalla lisääntyvät. Rajatummissa asetelmissa saadaan tarkempia tuloksia kapeammalta alueelta, jolloin pystytään paremmin kontrolloimaan erilaisia väliin tulevia muuttujia. Todellisessa maailmassa meno saattaa äityä vilkiksi ja olla varsin vaihtelevaa, jolloin siitä on vaikeaa sanoa mitään pitävää.

Tarkastelua voidaan laajentaa pelkästä tekstin lukemisesta tutkimusaineistojen analyysiin. Esimerkiksi historian tutkimuksessa informaatioyksiköiden syvälukeminen on eräs analyysimenetelmä, jossa voimakkaasti tulkitaan informaatioyksiköitä muiden informaatioyksiköiden antaman evidenssin varjolla ja aiemman ymmärryksen kautta. Tekstejä ei näin ollen käytetäkään vain niiden kirjoittajan aikomaan alkuperäiseen tarkoitukseen, vaan myös tekstin luomisen aikana vallinneiden kulttuuristen ja yhteiskunnallisten ilmiöiden tulkinnassa (esimerkiksi kirjeet tai kuvat kertomassa so- ta-ajasta ja sotarintaman elämästä).⁴⁴

Informaation synteesi ja raportointi

Monimutkaisissa tehtävissä pelkästään informaatioyksiköiden kanssa työskentely ei riitä, vaan tiedollisen työskentelyn tulokset täytyy usein myös jollain tavoin dokumentoida. Informaatioyksiköitä käytetään evidenssinä ja oikeutuksena esimerkiksi tieteellisessä tekstissä (esimerkiksi historian tutkijat).⁴⁵ Myös opiskelijat tuottavat opinnoissaan erilaisia tieteelliseen tietoon perustuvia tekstejä osana akateemisten taitojen kehittymistä. Tässä aktiviteetissa toimija uudelleen asettelee ja tiivistää informaatioyksiköiden

43 Ks. esim. Kumpulainen & Järvelin 2010.

44 Kumpulainen ym. 2020.

45 Ks. esim. Korkeamäki & Kumpulainen 2019.

pääkohdat.⁴⁶ Useasta eri dokumentista ja informaatioyksiköstä yhdistellyn tiedon tuottaminen on haastavampaa kuin yhden dokumentin tai tekstin referointi. Toimijan täytyy muodostaa yhtenäinen käsitys ja ymmärrys siitä, kuinka muuttaa informaatiosisällöt uuteen rakenteeseen ja sisällyttää tekstiinsä jopa toisilleen vastakkaisia näkemyksiä aiheesta.

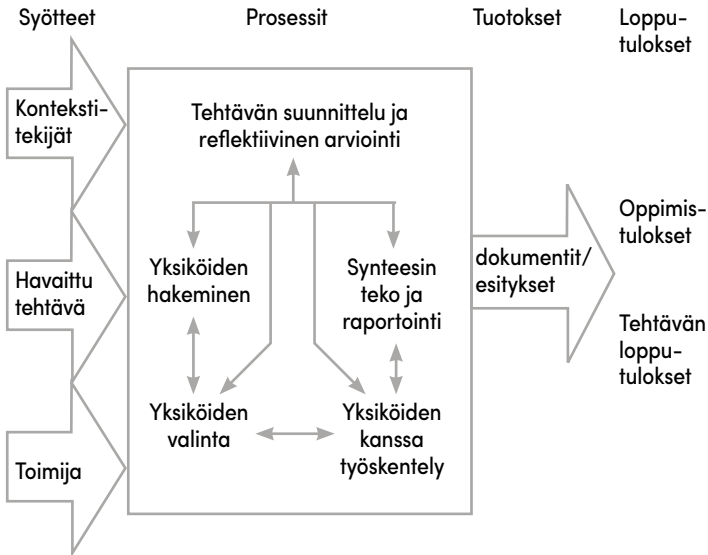
Ohjelmateorian puitteissa esitettynä informaation synteesi ja raportointi ottaa syötteenä prosesseihinsa pääasiassa samankaltaisia syötteitä kuin informaatioyksiköiden kanssa työskentelyssä (kontekstitekijät, havaitut tehtävätarpeet ja toimijan ominaisuudet), mutta työskenneltyään informaatioyksiköiden kanssa toimijan tiedollinen taso on uusi. Lisäksi tehtävän tarpeiden ymmärrys on muuttunut, koska informaation synteessin ja raportoinnin vaatimukset täytyy ottaa huomioon: mikä on tuotoksen lajityyppi, miten oletetun yleisön kanssa tulee keskustella ja niin edelleen. Uutena syötteenä ovat prosessoitu informaatio edellisestä aktiviteetista ja kirjoitusapuvälineet. Kirjoitusapuvälineinä voivat toimia esimerkiksi erilaiset tekstinkäsittelyohjelmistot, kuten MS Word, tai viittaustenhallintatyökalut, kuten Zotero tai Mendeley.

Prosesseina tässä aktiviteetissa ovat suunnittelu, kääntäminen ja tarkastelu.⁴⁷ Suunnittelussa toimija organisoii ajatuksensa tehtävän lopputuloksesta perustuen aiempaan tietämykseensä ja jo hankittuihin informaatioyksiköihin. Kääntäminen tarkoittaa ajatusten muuttamista kielelliseen esitysmuotoon muokkaamalla esimerkiksi luonnoksia ja esityksiä suunnitelman mukaisesti. Tarkastelussa toimija lukee luonnoksia havainnoidakseen ongelmia tekstissä tai sen rakenteissa. Tästä syntyy prosessin tuotos eli uusi dokumentti tai jonkinlainen esitys. Lopputuloksina syntyy yksilöllistä eli käsitteellistä ja faktuaalista osaamista sekä varsinaisen tehtävän lopputulos, kuten opiskelijan essee. Oppimista ja tehtävän lopputulosta, esimer-

46 Sormunen & Lehtiö 2011.

47 Vrt. Flower & Hayes 1981.

2.3 Tehtävälähtöinen informaatiovuorovaikutus



Kuvio 3. Tehtävälähtöisen informaatiovuorovaikutuksen evaluointikehikko.

kiksi opiskelijan tuottamaan esseenä, voidaan arvioida sen laadun perusteella.

Tehtävälähtöisen informaatiovuorovaikutuksen mallia hyödyn-
tävissä tutkimuksissa on yksittäisten aktiviteettien suunnitteluprosessi katsottu osaksi koko laajemman tehtävän suunnittelun ja arvi-
oinnin aktiviteettia.⁴⁸ Alkuperäisessä mallissa asiaa ei kuvattu aivan
näin, mutta jokaisella aktiviteetilla on vuorovaikutussuhde tehtävän
suunnitteluun ja reflektiiviseen arviointiin. Kun tarkastellaan kaik-
kia TBII-mallin aktiviteetteja yhdessä, on käytännössä selkeämpää
ottaa tämänkin yksittäisen aktiviteetin suunnitteluprosessi osaksi
tehtävän suunnittelua ja reflektiivistä arviointia, jolloin kaikkien ak-
tivateettien suunnitteluprosessit ovat tarkasteltavissa ja vertailtavissa
yhdessä ja rinnakkain.

48 Esim. Korkeamäki & Kumpulainen 2019; Kumpulainen ym. 2020;
Kumpulainen & Late 2022; Late & Kumpulainen 2021.

Tehtävälähtöisen informaatiovuoro- vaikutuksen malli

Tehtävälähtöisen informaatiovuorovaikutuksen kehikon yksittäisiä aktiviteetteja voidaan hyödyntää informaatiovuorovaikutuksen ja siihen liittyvien ilmiöiden tarkastelussa, mutta myös yli kaikkien edellä esitettyjen yksittäisten aktiviteettien. Kuvio 3 esittää yksittäiset aktiviteetit ohjelmateorian näkökulmasta: syötteenään tehtävälähtöinen informaatiovuorovaikutus saa kontekstitekijät, havaitun tehtävän sekä toimijan. Prosesseina ovat yksittäiset aktiviteetit: tehtävän suunnittelu ja reflektiivinen arviointi, informaatioyksiköiden hakeminen, informaatioyksiköiden valinta, informaatioyksiköiden kanssa työskentely sekä informaation synteesi ja raportointi. Kuviossa on tilan säästämisen vuoksi käytetty termiä yksikkö termin tietoyksikkö sijasta. Näiden kaikkien aktiviteettien välillä on molemmin suuntaista vuorovaikutusta, ja järjestys osasten välillä voi vaihdella.

Arviointikehikon vahvuuksia on useita. Ensinnäkin se kattaa koko tehtävälähtöisen informaatiovuorovaikutuksen kaaren suunnittelusta tehtävän lopputulokseen kaikkine edellä lueteltuine aktiviteetteineen.

Toiseksi se tarjoaa tukea aktiviteetteihin perustuvalla evaluoinnille eli inhimillisen toiminnan lähtökohdista tapahtuvan informaatiovuorovaikutuksen tarkastelulle. Yksittäisten aktiviteettien osalta tarjotaan ohjelmateoriat, joista voidaan valita sopivat kokonaisuudet tai vaihtoehtoisesti tarkastella koko aktiviteettikaarta. Kehikko muistuttaa tutkijaa ja arvioijaa siitä, että lopputuloksiin vaikuttaa useita eri aktiviteetteja, ja se auttaa ymmärtämään esimerkiksi yksittäisen tietoresurssin evaluointia laajemmassa yhteydessä.

Kolmanneksi kehikko tukee sekä syklisen että vaiheittaisen tehtäväprosessin analysointia, mikä tukee hyvinkin erilaisten tehtävien suoritusprosessien tarkastelua. Yksinkertaiset tehtävät saattavat olla varsin suoraviivaisia, mutta monimutkaiset, kompleksiset tehtävät vaativat iteratiivisia ja vaiheittaisia ja eri tavoin jäsentyneitä suoritustapoja. Suoritustavat saattavat sisältää eri tavoin järjestyneitä ak-

tiviteettiprosesseja, jotka taas niveltyvät toisiinsa ja tapahtuvat yhtä aikaa.

Neljäntenä vahvuutena on se, että kehikko keskittyy informaatiovuorovaikutuksen arviointiin eli ihmisen toimintaan ja erilaisen työkalujen avulla saavutettuihin hyötyihin tehtävän suorituksessa. Toiminnan keskiössä ei siis ole mikään työkalu tai tietokanta, vaan se, miten niitä hyödynnetään jonkin tavoitteellisen toiminnan apuna.

Viidentenä vahvuutena kehikossa on siinä vaikuttavien prosessin osasten sekä näiden tuottamien vaikutusten huomioiminen. Siinä määritellään syy-yhteysjärjestelmiä aktiviteettien keskeisten osasten tuotoksille ja lopputuloksille tuottamalle hyödyllä – tai harmille. Siten kehikko auttaa suuntaamaan huomiota keskeisiin osasiin ja jättämään toissijaiset vähemmälle huomiolle. Toisin sanoen se auttaa ymmärtämään, mikä on oikeasti tärkeää, ja estää yhdentekevien nyanssien tarkastelun.

Kuudenneksi kehikko sopii monien erilaisten informaatiotyökalujen ja -välineiden evaluointiin. Erityyppiset tehtävät vaativat erilaisia työkaluja, ja vaikka vielä ei ole olemassa välineitä, jotka tukisivat koko TBII-prosessia, sellaiset ovat kuitenkin tulevaisuudessa mahdollisia.

Mallilla on kuitenkin myös heikkoutensa. Kehikko on teoreettinen, ja sen sovellettavuus empiriaan eli varsinaisten tutkimuskohdeiden havainnointiin oli sen kehittämisaikana jokseenkin arvailujen varassa. Kehikko perustui tutkijoiden aiempiin kokemuksiin tarkastelluista ilmiöistä, eikä sitä silloin testattu. Lisäksi se on luotu tietynlainen oppimistehtävä mielessä, eikä sen sovellettavuudesta muihin tehtäviin välttämättä ollut mitään todistetta. On kuitenkin totta, että hyvin usein haastavat, kompleksiset tehtävät sisältävät lukuisia oppimistehtävälle tyypillisiä ominaisuuksia. Siten mallin väitetään sopivan hyvin muihinkin tehtäviin, mikä pitääkin nykytiedon valossa paikkansa. Myöhemmin empiirisissä tutkimuksissa on havaittu, että esimerkiksi tietoyksiköiden hakeminen ja tietoyksiköiden valitseminen voivat olla käytännössä esimerkiksi haastattelututkimuksissa

mahdottomia erottaa toisistaan.⁴⁹ Kuitenkin näiden kahden aktiviteetin erillinen tarkastelu voi olla mahdollista esimerkiksi vain näitä kahta aktiviteettia tarkastelevissa suppeammissa, kokeellisissa tutkimusasetelmissä.

Puutteina voidaan mainita myös, että alkuperäisen kehikon puitteissa kontekstin käsittely on hankalaa. Kehikko ei tarjoa selitystä tai jäsennystä sille, miten konteksti tai kontekstitekijät vaikuttavat. Esimerkiksi se, että dokumenteilla on myös oma kontekstinsa ja syntyhistoriansa, ei vain toimijoilla, jää huomioitta. Konteksti ei ole vain yksi asia, jonka voi ottaa syötteenä prosessiin, kuten organisaation asettamat rajoitteet, vaan se saattaa olla monitahoinen ja hankalasti jäsennettävä kokonaisuus. Se vaikuttaa monin eri tavoin aktiviteettien eri osasiin. Kontekstin määrittely on jäänyt Kumpulaisen ja Laten mukaan epämääräiseksi, eikä konteksti välttämättä ole hahmotettavana yksittäisinä syöteinä eri aktiviteetteihin. Heidän muunnellussa versiossaan kehikosta aktiviteetit olivat upotettuina kontekstien eri kerrostumiin, ja nämä erilaiset ilmentymät kontekstista vaikuttivat siihen, millaisia esteitä informaatiovuorovaikutuksen aikana kohdattiin.⁵⁰

Rajoituksistaan huolimatta kehikko on varsin käyttökelpoinen laajojen informaatiovuorovaikutusprosessien tarkasteluissa. Lisäksi se yleisenä kehikkona tuntuu suorastaan houkuttelevan sen soveltamiseen ja edelleen kehittelyyn eri aihealueilla tai toimintaympäristöissä, joissa on omia erityisvaatimuksiaan joko informaatioprosessien, -välineiden tai -resurssien suhteen. Se sitoo inhimillisen informaatiotoiminnan näkökulmasta tietoresurssit ja informaatiotyökalut siihen, mitä toiminnalla tavoitellaan ja mahdollistaa pelkkien tuotosten tarkastelun lisäksi pidemmän tähtäimen lopputulosten huomioimisen. Kehikko avaa uusia mahdollisia näkökulmia myös tiedonhaun, tiedonhankinnan ja informaatiokäyttäytymisen tarkasteluun.

49 Korkeamäki & Kumpulainen 2019; Late & Kumpulainen 2021.

50 Kumpulainen & Late 2022.

Tietoa siitä, miten ihmiset hyödyntävät tietojärjestelmiä ja erilaisia tietoresursseja tehtävissään tai miten nämä tietoresurssit tukevat tehtäviä, on erittäin vähän. Tieto on pirstaleista, ja monet tutkimukset tarkastelevat vain pieniä osasia tästä ilmiökentästä. Näistä pienistä osasista kasvaa ymmärrys ilmiöstä yli yksittäisten tutkimusten. Tämän ymmärryksen avulla voidaan luoda täysin uudenlaisia tapoja tukea ihmisiä erilaisilla tietoresursseilla ja informaatioprosesseilla tehtävien suorituksessa. Erilaisissa ympäristöissä ja tilanteissa sekä erilaisten tietoresurssien kanssa toimimisesta olisi siksi hyvä saada lisää tutkittua tietoa.

Lähteet

- Arguello, Jaime & Wan-Ching Wu & Diane Kelly & Ashley Edwards (2012) Task complexity, vertical display and user interaction in aggregated search. *Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval*, 435–444. <https://doi.org/10.1145/2348283.2348343>
- Arvola, Paavo & Jaana Kekäläinen (2010) Simulating user interaction in result document browsing. Teoksessa Leif Azzopardi (toim.) *Workshop proceedings of the 33rd annual international ACM SIGIR conference: Simulation of interaction, automated evaluation of interactive IR*. Geneve: ACM, 27–28.
- Belkin, Nicholas J. & Robert N. Oddy & Helen M. Brooks (1982) Ask for information retrieval: Part I. Background and theory. *Journal of documentation*, 38:2, 61–71.
- Blandford, Ann & Simon Attfield (2010) *Interacting with information*. Morgan & Claypool Publishers.
- Byström, Katriina & Kalervo Järvelin (1995) Task complexity affects information seeking and use. *Information processing & management*, 31:2, 191–213.
- Byström, Katriina & Sanna Kumpulainen (2020) Vertical and horizontal relationships amongst task-based information needs. *Information processing & management*, 57:2, 102065. <https://doi.org/10.1016/J.IPM.2019.102065>
- Fidel, Raya (2012) *Human Information Interaction: An ecological approach to information behavior*. Cambridge, MA: MIT Press.
- Flower, Linda & John R. Hayes (1981) A cognitive process theory of writing. *College composition and communication*, 32:4, 365–387.
- Fujii, Lee Ann (2017) *Interviewing in social science research: A relational approach*. New York: Routledge. <https://doi.org/10.4324/9780203756065>
- Huuskonen, Saila & Pertti Vakkari (2013) “I did it my way”: Social workers as secondary designers of a client information system. *Information processing & management*, 49:1, 380–391.

- Ingwersen, Peter (1992) *Information retrieval interaction*. Lontoo: Taylor Graham.
- Ingwersen, Peter & Järvelin Kalervo (2005) *The Turn: Integration of information seeking and retrieval in context (The Information retrieval series)*. Dordrecht: Springer.
- Järvelin, Kalervo & Pertti Vakkari & Paavo Arvola & Feza Baskaya & Anni Järvelin & Jaana Kekäläinen & Heikki Keskustalo & Sanna Kumpulainen & Miamaria Saastamoinen & Reijo Savolainen & Eero Sormunen (2015) Task-based information interaction evaluation: The viewpoint of program theory. *ACM transactions on information systems*, 33:1. <https://doi.org/10.1145/2699660>
- Kelly, Diane & Jaime Arguello & Ashlee Edwards & Wan-ching Wu (2015) Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. *ICTIR 2015 - Proceedings of the 2015 ACM SIGIR international conference on the theory of information retrieval*, 101–110. <https://doi.org/10.1145/2808194.2809465>
- Korkeamäki, Laura & Sanna Kumpulainen (2019) Interacting with digital documents: A real life study of historians' task processes, actions and goals. *CHIIR 2019 - Proceedings of the 2019 conference on human information interaction and retrieval*. <https://doi.org/10.1145/3295750.3298931>
- Krathwohl, David R. (2002) A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41:4, 212–218.
- Kuhlthau, Carol C. (1991) Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science*, 42:5, 361.
- Kuhlthau, Carol C. (1993) *Seeking meaning: A process approach to library and information services*. New York: Ablex Publishing Corporation.
- Kumpulainen, Sanna (2013) *Task-based information access in molecular medicine: Task performance, barriers, and searching within a heterogeneous information environment*. Väitöskirja. Tampere: Tampereen yliopisto. <http://urn.fi/URN:ISBN:978-951-44-9287-7>
- Kumpulainen, Sanna (2014) Trails across the heterogeneous information environment: Manual integration patterns of search systems in molecular medicine. *Journal of documentation*, 70:5, 856–877. <https://doi.org/10.1108/JD-06-2013-0082>
- Kumpulainen, Sanna (2017) Task-based information searching. Teoksessa John D. McDonald & Michael Levine-Clark (toim.) *Encyclopedia of library and information sciences*. Boca Raton: CRC Press, 4526–4536. <https://doi.org/https://doi.org/10.1081/E-ELIS4>
- Kumpulainen, Sanna & Kalervo Järvelin (2010) Information interaction in molecular medicine: Integrated use of multiple channels. *IliX 2010 - Proceedings of the 2010 information interaction in context symposium*. <https://doi.org/10.1145/1840784.1840800>
- Kumpulainen, Sanna & Kalervo Järvelin (2012) Barriers to task-based information access in molecular medicine. *Journal of the American society for information science and technology*, 63:1. <https://doi.org/10.1002/asi.21672>
- Kumpulainen, Sanna & Kalervo Järvelin & Sami Serola & Aiden R. Doherty & Alan F. Smeaton & Daragh Byrne & Gareth J. F. Jones (2009) Data collection methods for task-based information access in molecular

2.3 Tehtävälähtöinen informaatiovuorovaikutus

- medicine. *Proceedings of the 1st international workshop on mobilizing health information to support healthcare-related knowledge work, MobiHealthInf 2009 in conjunction with BIOSTEC 2009*.
- Kumpulainen, Sanna & Heikki Keskustalo & Boyang Zhang & Kostas Stefanidis (2020) Historical reasoning in authentic research tasks: Mapping cognitive and document spaces. *Journal of the association for information science and technology* 71:2, 230–241. <https://doi.org/10.1002/asi.24216>
- Kumpulainen, Sanna & Elina Late (2022) Struggling with digitized historical newspapers: Contextual barriers to information interaction in history research activities. *Journal of the association for information science and technology*, 73:7, 1012–1024. <https://doi.org/10.1002/ASI.24608>
- Late, Elina & Sanna Kumpulainen (2021) Interacting with digitised historical newspapers: understanding the use of digital surrogates as primary sources, *Journal of documentation*, 78:7, 106–124. <https://doi.org/10.1108/JD-04-2021-0078>
- Lin, Shin-jeng & Nicholas J. Belkin (2000) Modeling multiple information seeking episodes. *Proceedings of the ASIS annual meeting*, 37, 133–147.
- McDonald, Seonaidh (2005) Studying actions in context: A qualitative shadowing method for organizational research. *Qualitative research*, 5:4, 455–473.
- Preece, Jenny & Yvonne Rogers & Helen Sharp & David Benyon & Simon Holland & Tom Carey (1994) *Human-computer interaction*. Wokingham: Addison-Wesley Longman Ltd.
- Roos, Annikki & Sanna Kumpulainen & Kalervo Järvelin & Turid Hedlund (2008) The information environment of researchers in molecular medicine. *Information research*, 13:3.
- Rossi, Peter H. & Mark W. Lipsey & Howard E. Freeman (2003) *Evaluation: A systematic approach*. Sage Publications.
- Ruthven, Ian (2009) Interactive information retrieval. *Annual review of information science and technology*, 42:1, 43–91. <https://doi.org/10.1002/aris.2008.1440420109>
- Saastamoinen, Miamaria & Kalervo Järvelin (2017) Search task features in work tasks of varying types and complexity. *Journal of the association for information science and technology*, 68:5, 1111–1123. <https://doi.org/10.1002/asi.23766>
- Saastamoinen, Miamaria & Sanna Kumpulainen (2014) Expected and materialised information source use by municipal officials: Intertwining with task complexity. *Information research*, 19:4.
- Saastamoinen, Miamaria & Sanna Kumpulainen & Pertti Vakkari & Kalervo Järvelin (2013) Task complexity affects information use: A questionnaire study in city administration. *Information Research*, 18:4.
- Shah, Chirag & Chris Leeder (2015) Exploring collaborative work among graduate students through the C5 model of collaboration: A diary study. *Journal of information science*, 42:5, 609–629. <https://doi.org/10.1177/0165551515603322>
- Sormunen, Eero & Leeni Lehtiö (2011) Authoring Wikipedia articles as an information literacy assignment: Copy-pasting or expressing new

- understanding in one's own words? *Information Research*, 16:4. <https://eric.ed.gov/?id=EJ956120>
- Sun, Xu & Sarah Sharples & Stephann Makri (2011) A user-centred mobile diary study approach to understanding serendipity in information research. *Information research*, 16:3.
- Taylor, Robert S. (1968) Question-negotiation and information seeking in libraries. *College & research libraries*, 29:3, 178–194.
- Vakkari, Pertti (2001) A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of documentation*, 57:1, 44–60. <https://doi.org/10.1108/EUM0000000007075>
- Vakkari, Pertti & Salla Huuskonen (2012) Search effort degrades search output but improves task outcome. *Journal of the American society for information science and technology*, 63:4, 657–670. <https://doi.org/10.1002/asi.21683>
- Wildemuth, Barbara M. & Luanne Freund & Elaine G. Toms (2014) Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of documentation*, 70:6, 1118–1140.
- Wilson, Thomas D. (1981) On user studies and information needs. *Journal of documentation*, 37:1, 3–15. <https://doi.org/10.1108/ebo26702>

Informaatio ja tiedonhakujärjestelmät

3.1 Hakukoneista kielimalleihin

Heikki Keskustalo, Paavo Arvola ja Sanna Kumpulainen

- Hakukoneiden toiminta ja kehitys perustuvat tiedonhaun tieteelliseen tutkimukseen.
- Perinteiset hakukoneet hakevat dokumentteja ja niiden osia hakualgoritmeja hyödyntäen, jolloin keskeisiä kiinnostuksen kohteita ovat haetut informaatioyksiköt ja niiden väliset suhteet.
- Perinteisten hakukoneiden rinnalle ovat tulleet informaatiota *muodostavat* eli generoivat järjestelmät.
- Muuttuva teknologia ja sen käyttö tarjoavat alan tutkimukselle alati uusia haasteita.

Tietoa on haettu erilaisilla menetelmillä vuosisatojen ajan – jo kauan ennen nykyisen kaltaisia hakukoneita. Kirjallista materiaalia alkoi vähitellen kertyä erilaisiin kirjastoihin ja kokoelmiin niin paljon, että halutun tiedon löytämiseksi ei enää riittänyt se, että dokumentit, kuten kirjat, olivat fyysisesti saatavilla. Piti kehittää tehokkaampia menetelmiä, kuten kortistoihin perustuvia luetteloita, tiedon löytämiseksi laajoista kokoelmista. Tietotekniikan nopea kehitys 1950-luvulta alkaen ja erityisesti World Wide Webin tulo 1990-luvulla ovat mullistaneet tiedonhaun alan, ja nykyisin *tiedonhaku* on usein läsnä ihmisten arjessa ja työssä – ja samalla merkittävä tutkimusalue.

Tiedonhaun esittelyssä lähdetään usein liikkeelle kuvailemalla abstraktin hakukoneen rakenneosat ja dokumenttien tallennusprosessin ja hakumekanismin yleiset periaatteet. Myös tässä luvussa kuvaamme tiedonhakua tähän tapaan. Huomioimme kuitenkin samalla tehtävälähtöisen informaatiovuorovaikutuksen näkemyksen (ks. luku 2.3), jonka mukaan tiedonhaku tulisi ymmärtää osana

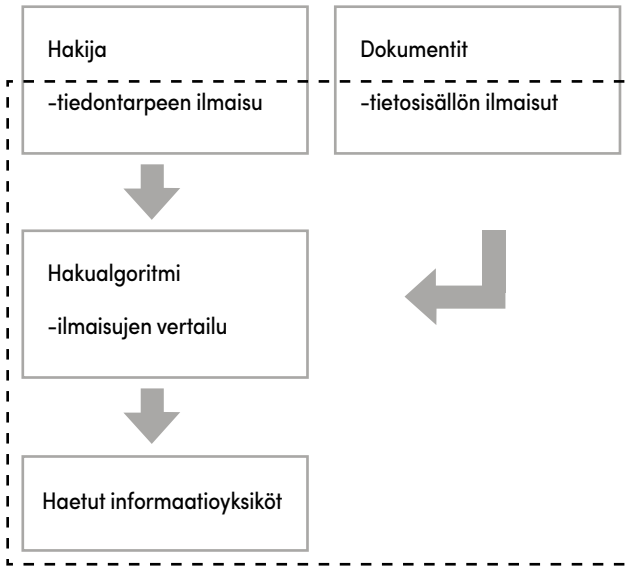
laajempaa ja monimuotoista vuorovaikutusta. Lähdemme liikkeelle *hakijan* suunnasta pyrkien korostamaan käytännön tiedonhakuun liittyvää moninaisuutta ja monenlaista kompleksisuutta. Hakijan luonnehdinnan ohella kuvaamme dokumenttien prosessoinnin, hakemisen ja tiedonhaun evaluoinnin pääideoita. Korostamme näkökulmaa, jonka mukaan tiedonhaku olisi hyvä ymmärtää osana laajempaa vuorovaikutusprosessia.¹

Hakujärjestelmiä kehitettäessä arviointi eli *evaluointi* on keskeistä. Evaluointi voi kohdistua moniin eri seikkoihin. Kohteena voi olla vaikkapa haun tuloksen laatu tai haun teknisen toteutuksen tehokkuus. Croft, Metzler ja Strohman² tarjoavat hyvän yleiskatsauksen hakukoneiden evaluointiin. Käsillä olevassa luvussa pyrimme osaltamme piirtämään kuvaa tiedonhaun tutkimuksen *mallipohjaisuudesta*. Mallit redusioivat monimuotoisen todellisuuden rajatuksi joukoksi käsitteellisiä piirteitä ja niiden välisiä suhteita. Tämä on mallien väistämätön heikkous ja vahvuus. Heikkoutena on suppea ja abstrahoitu näkemys tarkastellusta ilmiöstä ja sen todellisen monimuotoisuuden väheneminen. Toisaalta mallin vahvuutena on tarkastellun ilmiön erittely ja erottaminen ympäröivistä ilmiöistä sekä analyyttisen ja tarkkarajaisen tarkastelun mahdollistaminen. Annamme esimerkkejä tiedonhaun tutkimuksista, jotka kuvastavat erilaisia tapoja hahmottaa tiedonhakua ilmiönä ja erityisesti sitä, millaisena tiedon hakija tutkimuksissa näyttäytyy. Tiedonhaun peruskäsitteiden havainnollistamiseen soveltuu hyvin tiedonhaun abstrakti yleinen malli (kuvio 1). Malli auttaa osoittamaan ilmiön tutkimiseen liittyviä haasteita.

Kuviossa 1 esitetyn tiedonhaun yleisen mallin pääelementit ovat hakija, dokumentit, hakualgoritmi ja haetut informaatioyksiköt. *Hakijalla* tarkoitetaan hakujärjestelmän (täsmällisemmin ilmaistuna tiedon tallennus- ja hakujärjestelmän) käyttäjää. Kiinnostavia hakijan piirteitä ovat tyypillisesti hakijan tiedontarpeet ja tiedontar-

1 Järvelin ym. 2015.

2 Croft ym. 2010.



Kuvio 1. Tiedonhaun abstrakti yleinen malli, johon evaluoinnin kohteita on merkitty katkoviivalla.

peiden ilmaisu (usein sanojen avulla ilmaistut *kyselyt*). *Dokumentit* sisältävät hakujärjestelmän avulla haettavan tietosisällön. Ne voivat olla esimerkiksi verkkosivuja tai pikaviestipalvelun viestejä. Jotta tietoa voitaisiin hakea riittävän tehokkaasti, tiedon tallennus- ja hakujärjestelmä tallentaa dokumenttien tietosisällön sopiviksi tietorakenteiksi, joita hyödyntäen *informaatioyksiköiden* eli dokumenttien tai niiden osien löytäminen on mahdollista. Esimerkki tällaisesta aputietorakenteesta on hakujärjestelmän *indeksi*, joka ilmaisee dokumenttien sanat ja osoitetiedot siitä, missä dokumenteissa sana esiintyy.

Hakualgoritmi vastaanottaa hakijan antaman tiedontarpeen ilmaisuuden ja dokumenttien ilmaisu ja vertailee niitä. Vertailun perusteella voidaan hakea valittuja informaatioyksiköitä, jotka näytetään hakijalle käyttöliittymässä (esimerkiksi hakukoneen palauttamana otsikkolistauksena). Tiedon hakija puolestaan *arvioi* haun onnistu-

neisuutta, usein intuitiivisesti. Tutkija sen sijaan haluaa usein vertailla hakujärjestelmien tai -menetelmien keskinäistä paremmuutta numeerisen täsmällisesti. Tällöin evaluoinnin käsitettä joudutaan tarkentamaan. Evaluoinnin kohteeksi voidaan rajata valittu hakuprosessin elementti (ks. kuvion 1 katkoviiva), vaikkapa se, kuinka hyvin haettujen informaatioyksiköiden sisältö vastaa kyselyssä ilmaistuun aiheeseen. Tällainen *aiherelevanssiin* pohjautuva näkökulma on keskeinen, mutta kuitenkin vain yksi evaluoinnin monista mahdollisuuksista. Tarkastelemme kysymystä evaluoinnista yksityiskohtaisemmin luvussa 3.2.

Kuvion 1 tiedonhaun yleinen malli luonnehtii suurpiirteisesti tyypillistä ja yleistä hakutilannetta, jossa hakija pyrkii muodostamaan yhteyden relevantteihin dokumentteihin hyödyntäen apuna hakujärjestelmää. Malli havainnollistaa tilannetta korkealla abstraktiotasolla ja tarjoaa samalla lähtökohdan pohtia mallin neljää pääelementtiä ja tarkentaa niitä. Malli soveltuu myös tiedonhaun tutkimusten vertailemiseen ja luonnehtii sellaisenaan ytimekkäästi niin sanotun *testikokoelmapohjaisen* tutkimuksen pelkistettyä näkemystä tiedonhakutilanteesta.³ Mallista kuitenkin selvästikin puuttuu todellisen hakutilanteen piirteitä.

Hakija

Tiedonhaun yleisen mallin lähtökohtina olivat hakija ja dokumentit. Hakijalla on tiedontarpeita, joita hän pyrkii ilmaisemaan hakujärjestelmälle käyttäen apuna tyypillisesti hakusanoista koostuvaa kyselyä. Ihmisten tiedontarpeet ja heidän käyttämänsä järjestelmät voivat olla monentyyppisiä. Hakujärjestelmä hyödyntää *hakualgoritmia* relevanttien dokumenttien hakemiseen (ks. alaluku Hakualgoritmit). Hakutilanteeseen liittyy moninaisia haasteita. Hakijan näkökulmasta *kyselyn muotoileminen* on lähtökohtaisesti vaikeaa, sillä kysely on tyypillisesti hakutoimintaa motivoivan tarpeen suppea ja

3 Croft ym. 2010.

epätäydellinen kuvaus. Kyselyn kuitenkin tulisi kuvata haluttu informaation sisältö riittävän tyhjentävästi ja yksiselitteisesti. Muodostettu kysely voi olla monitulkintainen, sillä kyselyn ilmaukset saatavat viitata moniin eri käsitteisiin. Dokumentteja tarkasteltaessa vastaavana haasteena on sisällön ilmaisemiseen liittyvä *semanttinen avoimuus*: dokumentin tietosisältö on tyypillisesti ilmaistavissa monilla vaihtoehtoisilla tavoilla eri sanavalintoja käyttäen. Hakuaihetta erinomaisen hyvin luonnehtivat sanat saattavat esiintyä relevanttien ohella myös tyystin epärelevantteissa dokumenteissa. Lisäksi tiedontarpeen ilmaisemista vaikeuttaa se, että haluttu tieto ei ole välttämättä haettavissa suoraan, vaan hakija joutuu konstruoimaan tiedon *tulkiten* haettuja dokumentteja.⁴

Toimivien hakusanojen löytäminen ei ole suinkaan ainoa hakijan kohtaamista haasteista. *Tiedontarve* itsessään voi olla hakijalle epäselvä ja vaikeasti ilmaistava.⁵ Tiedontarve ja hakijan kyky ilmaista se voivat muuttua hakemisen aikana.⁶ Tiedonhaku toteutuu lisäksi usein *istuntona*, jonka kuluessa hakija muodostaa kyselyn, arvioi saamaansa tulosta ja yrittää tarvittaessa hakea uudelleen.⁷ Hakijat punnitsevat hakutuloksen lisäksi kyselyiden muodostamiseen ja tulosten tarkasteluun kuluva aikaa ja vaivaa. He voivat esimerkiksi suosia *lyhyitä* kyselyitä (mikäli mahdollista) sanavalintoja muokaten (esimerkiksi hakusanaa vaihtaen).⁸ Hakutuloksen tarkastelu voi keskittyä pelkästään haetun tuloslistan *alkupäähän*.⁹ Lisäksi hakijalla voi olla monia henkilökohtaisia ja tilannesidonnaisia kriteerejä sille, millainen hakuprosessi on ollut riittävän onnistunut. Esimerkiksi dokumentin tuoreus, subjektiivinen uutuusarvo ja ymmärrettävyys voivat olla keskeisiä kriteereitä aiheenmukaisuuden lisäksi. Toisaalta aivan muutkin hakuun liittyvät piirteet kuin *dokumentin relevanssi*

4 Vrt. Ingwersen & Järvelin 2005, 114–116.

5 Belkin 1980.

6 Kuhlthau 1991; Vakkari 2000.

7 Jansen ym. 2000.

8 Jansen ym. 2000; Vakkari 2002; Ruthven 2008; Vakkari 2000.

9 Jansen ym. 2000; Ruthven 2008; Azzopardi 2007.

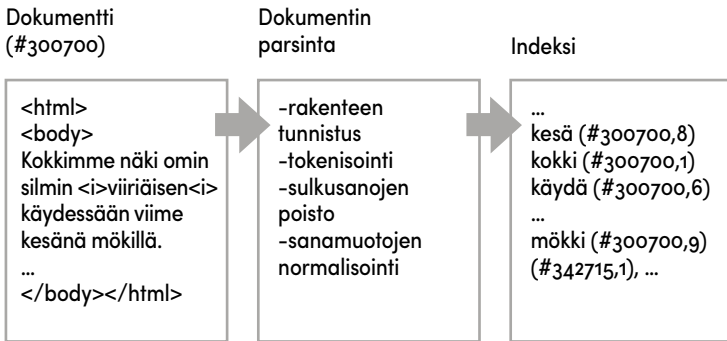
– esimerkiksi hakemisen vaivattomuus ja nopeus – voivat olla hakijalle jopa ratkaisevan tärkeitä.

Yhteenvetona voidaan todeta, että ihmiset hakevat dokumentteja monenlaisissa tilanteissa ja erilaiset tilannetekijät, kuten kiire, voivat vaikuttaa esimerkiksi siihen, montako ja minkä tyyppistä informaatioyksikköä hakija tarvitsee, mitä ”relevantit dokumentit” hakijalle tarkoittavat ja paljonko hakemiseen voidaan käyttää aikaa ja vaivaa. Tämän tyyppisiä tilannetekijöitä voidaan pyrkiä huomioimaan tiedonhaun tutkimuksissa eri tavoin. Tiedonhaun tutkimuksia voidaan yleisesti ottaen ajatella todellisten hakutilanteiden *simulatioina*, joissa jäljitellään yksinkertaistettua tilannetta. Yksittäisessä tutkimuksessa käytetty hakijan ja hakuprosessin malli voi olla tällöin karkean yksinkertainen tai hienojakoisempi ja monimutkaisempi. Ääritapauksessa tiedonhaun tutkija voi *pelkistää* hakijan hakujärjestelmälle annettujen kyselyiden muodostamaksi syötteen ja kyselyiden avulla haettujen dokumenttien aihelevanssin arvioimiseksi. Vastakkaisessa ääritapauksessa tiedonhaun tutkija seuraa todellisia tiedonhakijoita heidän suorittaessaan esimerkiksi työtehtäviä *autenttisissa* tilanteissa. Valaisemme asiaa esimerkkien avulla kirjan luvussa 2.3.

Dokumentit

Digitaaliset dokumentit voivat olla hyvin monentyyppisiä, esimerkiksi lehtiartikkeleita tai pikaviestejä, ja ne sisältävät tyypillisesti tekstiä, kuvia, ääntä ja videota. Kuvaamme seuraavassa lyhyesti dokumenttien prosessoinnin yleisperiaatetta hakujärjestelmän prosessoidessa haun kohteena olevia dokumentteja haettavaan muotoon. Käsitlemme prosessin yleisperiaatetta *tekstimuotoisen* tiedon tallennuksen ja haun näkökulmasta.

Dokumentin tekstisisältö voi yksinkertaisimmillaan koostua pelkästä sanojen jonosta vailla monimutkaisempaa rakennetietoa tai olla *rakenteinen* sisältäen erityyppisiä elementtejä, kuten *otsikon*, *leipätekstin* ja *kursivoinnin*. Rakennetieto koodataan tavallisesti do-



Kuvio 2. Hakukoneen suorittama dokumentin parsintaprosessi ja lopputuloksena syntynyt indeksi.

kumenttiin käyttäen sopivaa kuvauskieltä, kuten HTML tai XML.¹⁰ Tämän luvun alussa esitetyn yleisen mallin näkökulmasta olennaista on ymmärtää, että tietoa ei haeta hakutilanteessa ”suoraan” haettavien dokumenttien tekstielementtejä selaamalla, vaan hakuvaiheessa hakujärjestelmä hyödyntää rakentamaansa *indeksiä*. Indeksi on aputietorakenne, jonka perusidea on tuttu painettujen kirjojen loppulehdiltä. Tällaisissa indekseissä on listattu teoksen sisältämiä sanoja (yleensä aakkosjärjestyksessä) sivunumeroineen. Lukija voi indeksia käyttäen siirtyä suoraan sivunumerolle, jolla sana esiintyy, eikä hänen tarvitse selata kirjaa kokonaisuudessaan. Hakujärjestelmät hyödyntävät samaa periaatetta ja rakentavat dokumenteista *indeksin*, joka ilmaisee, missä sanat sijaitsevat (kuvio 2).

Kuviossa 2 on kuvattu yksittäisen dokumentin käsittelyä indeksin muodostamiseksi. Prosessi toteutetaan automaattisesti käyttäen tietokonealgoritmia, joka käy läpi hakujärjestelmään tallennettujen dokumenttien tekstit. Kuviossa vasemmalla on esitetty dokumentti, jonka varsinainen teksti on sijoitettu rakennetietoa koodaavien tunnisteiden `<body>` ja `</body>` väliin. Lisäksi yksittäinen sana *viiriäisen* on merkitty kursivoiduksi (*italics*) käyttäen tunnisteita `<i>` ja `</i>`.

¹⁰ Croft ym. 2010.

i>. Kuviossa keskellä on kuvattu indeksin rakentamisessa hyödynnettäviä vaiheita eli *dokumentin parsinta*. Parsinnan päävaiheet ovat *rakenteen tunnistus*, *tokenisointi*, *sulkusanojen käsittely* ja *sananmuotojen normalisointi*. Parsinnan lopputuloksena syntyy *indeksi*, jota tarkastelemalla on mahdollista vastata hakijan antamaan kyselyyn. Parsinnan yksityiskohdat vaikuttavat olennaisesti hakukoneen toimintaan ja käyttömahdollisuuksiin.

Rakenteen tunnistuksessa dokumentista tunnistetaan rakenneosia, kuten otsikkokenttiä tai muita elementtejä, jotka halutaan valikoida tarkempaa käsittelyä varten. Kuvion 2 esimerkissä dokumentista kerätään tekstimuotoinen informaatio. Prosessin alkaessa dokumentin tekstisisältö voidaan mieltää yhdeksi merkkijonoksi (esimerkiksi jonoksi UTF-8-merkistön merkkejä). Merkkijono sisältää tyypillisesti kirjaimia, numeroita, välilyöntejä ja välimerkkejä. Koska tiedonhaku tehdään käyttäen sanoja, *tokenisoinnissa* merkkijono ositetaan sanoiksi. Sanan käsitteen määrittely on epätriviaalia, mutta yksinkertaisimmillaan sanoja ovat peräkkäisistä kirjaimista tai numeroista koostuvat merkkijonot. Välilyönnit ja välimerkit auttavat osittamaan dokumentin sanoiksi. *Sulkusanojen poistossa* tietosisällön kannalta hyödyttömät sanat, esimerkiksi yleisimmät verbit ja partikkelit, poistetaan. Hyödyttömyyden määritteleminen on kuitenkin usein vaikeaa. Lopuksi sananmuotojen *normalisoinnissa* taipuneet sanamuodot voidaan muuntaa perusmuotoon ja esimerkiksi yhdyssanat jakaa osasanoihinsa, jotka viedään indeksiin. Yhdyssanojen ositus on haun näkökulmasta hyödyllistä, sillä hakijan viittaaman käsitteen (esimerkiksi [*viiriäinen*]) ilmaisut voivat esiintyä tekstissä paitsi taipuneina (*viiriäisen*, *viiriäiset* ja niin edelleen) myös yhdyssanan osana (esimerkiksi *viiriäisrehu*).

Parsinnan päätteeksi syntyneet sanamuodot tallennetaan indeksiin. Valmis indeksi pitää sisällään siis dokumenttien normalisoidut sanat osoitetietoineen. Kun *hakija* kirjoittaa kyselyn, järjestelmä etsii kyselyn sanoja indeksistä. Tietokonealgoritmi pystyy päättelämään kyselyyn täsmäävät dokumentit indeksin tietoja tarkastelemalla. Kuvion 2 esimerkki valaisee asiaa: indeksiä (kuviossa oi-

kealla) tarkastelemalla huomaamme, että sana ”kesä” (normalisoitu perusmuoto) esiintyy dokumentin #300700 kahdeksantena sanana (dokumentissa taivutusmuodossa ”kesänä”). Indeksien sanojen osoittelijat ovat käytännössä erittäin pitkiä.

Parsinnan eri vaiheet voidaan toteuttaa yksityiskohtien tasolla monin tavoin. Yksityiskohtien valintaan – esimerkiksi siihen, kuinka tokenisaatio tai normalisointi kannattaisi toteuttaa – vaikuttavat oletettu dokumenttien käyttö sekä dokumenttien ominaisuudet, kuten kieli ja dokumenttien tyyppi. Esimerkiksi lyhenteitä sisältävät tekstiviestit ja luonnollista kieltä sisältävät sanomalehtiartikkelit voitaisiin haluta tokenisoida hieman toisistaan eroavin säännöin. *Luonnollinen kieli* viittaa tässä ihmisten arjessa käyttämiin kieliin (esimerkiksi suomi, ruotsi ja englanti). Luonnollinen kieli tarjoaa lukuisia haasteita tiedonhauille. Tarkentavia tietoja dokumenttien parsinnan ja luonnollisen kielen aihepiireihin tarjoavat muun muassa Croft, Metzler ja Strohman¹¹ ja Alkula.¹²

Dokumenttien prosessointia kuvattiin edellä hakukoneen näkökulmasta. Dokumentteja voidaan pohtia myös dokumentin *kirjoittajan* näkökulmasta. Esimerkiksi verkkosivun laatija voisi haluta parantaa sivun näkyvyyttä hakutuloksissa. Laatija arvioi, että sivun sisällöstä kiinnostuneet hakijat voivat käyttää sivun sisältöön liittyviä hakusanoja. Laatija voi olla kiinnostunut ohjaamaan sivulle laajemminkin sellaisia hakijoita, jotka kuuluvat sivun kohderyhmään. Sivun sijoittumista voidaankin yrittää tietoisesti edesauttaa dokumenttia laadittaessa. Tällöin puhutaan *hakukoneoptimoinnista*. Optimointi voidaan jakaa toivottuun (*white hat*) ja kiellettyyn tai kyseenalaiseen (*black hat*). Ensin mainitussa optimoinnissa relevanttien sivujen sijoittumista autetaan tavoilla, jotka tukevat hakijan tarpeita, hakematta huomiota kyseenalaisin keinoin. Jälkimmäisessä optimointitavassa pyritään – esimerkiksi kaupallisten motiivien vuoksi – ohjaamaan ihmisiä sivuille, jotka eivät vas-

¹¹ Croft ym. 2010.

¹² Alkula 2000.

taa heidän odotuksiaan ja voivat olla jopa haitallisia. Optimoinnin menetelmiä on lukuisia, ja niiden toimivuus vaihtelee riippuen hakukoneen toimintaperiaatteesta. Esimerkkinä voidaan käyttää hakusanoja, joiden runsas esiintyminen relevantilla sivustolla on ymmärrettävää, mutta liioitellun tiheä esiintyminen voi muuttua epäilyttäväksi. Optimoiden (erityisesti kiellettyjen menetelmien käyttäjien) ja hakukoneiden kehittäjien suhdetta voidaan kuvata jatkuvana kilpajuoksuna. Tästä ovat kirjoittaneet muun muassa Duk, Bjelobrk & Carapina.¹³

Hakualgoritmit: kyselyt, täsmäytys ja järjestäminen

Hakukoneen tarkoitus on hakea, järjestää ja esittää hakijalle tietoa. Kansanomaisesti ilmaistuna tiedon hakeminen ja järjestäminen tapahtuu hakualgoritmillä, jonka muodostama tulos esitetään käyttäjälle käyttöliittymässä. Tuttu esimerkki tällaisesta on esimerkiksi Googlen tulossivu, jossa linkit parhaisiin haettuihin dokumentteihin esitetään sivulla ylimpänä.

Hakijan esittämän kyselyn lisäksi nykyaikainen internetin hakukone voi käyttää hyväkseen monenlaista muutakin informaatiota, jota tiedon hakijasta ja hänen toiminnastaan on teknisesti (evästeet ja IP-osoitteet) ja laillisesti saatavilla. Kuitenkin hakualgoritmin toiminnan *lähtökohta* on usein – ja joskus pelkästään – hakijan esittämä kysely, joka tyypillisesti koostuu yhdestä tai useammasta hakusanasta. Kyselyssä hakusanat on erotettu toisistaan välilyönnein tai (muilla) operaattoreilla, kuten JA (AND) ja TAI (OR).

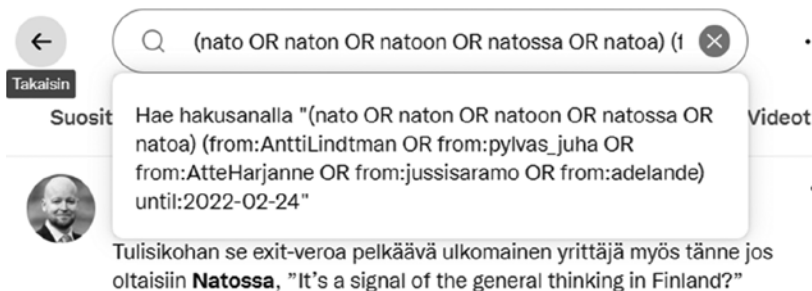
Täystäsmäytys

Usein hakualgoritmit toimivat siten, että hakusanojen lisääminen vähentää voimakkaasti hakutuloksien määrää. Esimerkiksi Google

¹³ Duk ym. 2013.

toimii juuri tällä tavoin. Tämä johtuu siitä, että teknisesti hakukone tulkitsee hakusanojen välillä olevan välilyönnin AND-operaattorin tapaan eli konjunktiksi. Toisin sanoen tullakseen valituksi tulosjoukkoon kaikkien hakusanojen on esiinnyttävä dokumentissa. Mikäli hakijalle riittää, että mikä tahansa hakusanoista esiintyy dokumentissa, on sanojen välissä käytettävä OR-operaattoria (disjunktio), mikäli tällainen on hakukoneessa käytettävissä. Lähes kaikki hakukoneet tukevat yleisten loogisten operaattoreiden AND, OR ja NOT käyttöä hakutuloksen kokoon vaikuttamiseksi. Hakukone vertaa käyttäjän esittämää kyselyä dokumentteihin edellä kuvattua indeksia hyödyntäen. Tätä vertailua kutsutaan täsmäyttämiseksi (*matching*), minkä voi katsoa olevan hakukoneen toiminnan ydin.

Jos hakualgoritmi valitsee halutulokseen *pelkästään* hakulausekkeeseen täysin täsmäyvät dokumentit, kuten kuviossa 3 hakuehdot täyttävät viestit, on kyseessä niin sanottu *täystäsmäytys*. Kuviossa esitetty hakukone tarjoaa loogisten operaattoreiden AND, OR ja NOT ohella muitakin operaattoreita. Kuitenkaan tätä kirjoitettaessa niin sanottu *wildcard*-operaattori *nato**, joka täsmäisi kaikkiin nato-alkuisiin sanoihin, ei ole käytössä. Sellaisen operaattorin avulla saataisiin kaikki sijamuodot ja yhdyssanat täsmäämään kyselyyn. Täystäsmäytys etsii annetun ehdon täyttävät dokumentit, muttei määrittele niille keskinäistä järjestystä. Käytännössä ehdon täyttävien dokumenttien esittämisjärjestys voisi perustua esimerkiksi julkaisuaikaan (muun muassa X), suosioon tai aakkosjärjestykseen. Internetin yleishakukoneet käyttävät yhtenä järjestyskriteerinään sivujen välisiä linkkirakenteita, jolloin sivut tai sivustot ikään kuin äänestävät toistensa luotettavuutta tai auktoriteettia hyperlinkkien (URL) avulla. Esimerkiksi Googlen perustajien lanseeraamassa PageRank-algoritmissa internetin sivut saavat linkityksiensä kautta (kyselyistä riippumattoman) pistemäärän, jonka avulla kyselyyn täsmäyvät sivut voidaan asettaa keskinäiseen järjestykseen.¹⁴



Kuvio 3. Näytönkaappaus hakulauskeista X-viestipalvelussa, jolla haetaan Natoa koskevia hallituspuolueiden eduskuntaryhmien puheenjohtajien kirjoittamia pikaviestejä ennen Venäjän hyökkäystä Ukrainaan.

Osittaistäsmäytys

Täystäsmäytyksessä hakija voi kontrolloida hakutuloksen kokoa ja – hakukoneen muista ominaisuuksista riippuen – mahdollisesti myös järjestystä. Peruskäyttäjälle loogisia operaattoreita AND, OR ja NOT sisältävät kyselyt eivät ole välttämättä luontevia ja helppokäyttöisiä, eikä täystäsmäytyksessä tulosten keskinäinen järjestys perustu dokumenttien täsmävytyteen. Siksi hakukoneiden tutkimuksessa ja kehityksessä onkin tiedonhaun tutkimuksen alku metreiltä lähtien kehitetty menetelmiä, joiden avulla hakijan hakusanoista tehdään päätelmiä tulosten järjestämiseksi automaattisesti. Menetelmiä, jotka järjestävät tulokset paremmuusjärjestykseen niiden täsmävytyden mukaan, kutsutaan *osittaistäsmäyttäviksi* menetelmiksi.

Osittaistäsmäytys pyrkii siihen, että tuloslistan kärjessä olevat dokumentit eli kyselyyn täsmävimmit dokumentit ovat todennäköisimmin hakijalle relevantteja. Osittaistäsmäytyksen yleisperiaatteena on, että kullekin dokumentille lasketaan algoritmisesti täsmävyysluku käyttäjän antaman kyselyn suhteen. Yksinkertainen esimerkki tällaisesta on termifrekvenssin (*term frequency*) hyödyntäminen sen ilmaistessa, montako kertaa sana toistuu kussakin dokumentissa. Esimerkiksi kyselyn sanan esiintyessä dokumentissa

A kolme kertaa ja dokumentissa B kaksitoista kertaa voitaisiin yksinkertaisen algoritmin avulla päätellä, että dokumentti B täsmää paremmin kyselyyn, ja se esitettäisiin hakutuloksissa ylempänä. Mikäli hakusanoja on kyselyssä useita, niiden esiintyvyydet yhdistetään sopivalla laskuoperaatiolla. Kertolasku olisi yksinkertainen lähtökohta laskuoperaatioksi ja sikäli intuitiivinen, että kyselyn sanan puuttuessa tietystä dokumentista (termin frekvenssi on o) muodostuisi dokumentin täsmäävyyslukuksi nolla, mikä vastaisi täystäsmäytyksen konjunkttiivista tulkintaa (dokumenttia ei haeta, jos jokin kyselyn sanoista puuttuu).

Termifrekvenssin käytön antama järjestys ei monen hakusanan kyselyssä olisi käytännössä riittävän hyvä, sillä kyselyn sanat ovat usein erottelukyvyltään keskenään eriarvoisia. Esimerkiksi kyselyssä *Viiriäisen muna* ensimmäinen sana (*viiriäisen*) on jälkimmäistä harvinaisempi ja sen esiintyminen ja toistuvuus dokumentissa ennustanee dokumentin relevanssia jälkimmäistä sanaa paremmin. Hakualgoritmissa, joka perustuu niin sanotulle TF.IDF-kaavalle (termifrekvenssi kertaa käänteinen dokumenttifrekvenssi), hakualgoritmi laskeekin kullekin hakusanalle myös harvinaisuusarvon IDF. Sanan dokumenttifrekvenssi on sitä suurempi, mitä useammissa kokoelman dokumenteissa sana esiintyy. Käänteinen dokumenttifrekvenssi kuvaa tätä ilmiötä kääntäen eli saa suurimpia lukuarvoja kokoelman harvinaisimmille sanoille.¹⁵ Sanan harvinaisuutta korostava lukuarvo voidaan laskea automaattisesti sanan esiintyvyydestietoja hyödyntäen, mutta jotkin hakukoneet antavat hakijalle lisäksi mahdollisuuden painottaa hakusanoja myös itse.

Termifrekvenssin ja käänteisen dokumenttifrekvenssin lisäksi dokumentin täsmäävyyslukuun voivat hakualgoritmeissa vaikuttaa muutkin tekijät, kuten se, kuinka tiheästi hakusanat esiintyvät (dokumentin pituusnormalisointi), kuinka lähekkäin ne esiintyvät tai missä ne esiintyvät (esimerkiksi dokumentin otsikossa tai linkeissä). Sofistikoituneemmat osittaistäsmäyttävät menetelmät voidaan

15 Spärck Jones 1972.

jaotella ainakin algebraan tai todennäköisyysteoriaan perustuviin täsmäytysmenetelmiin.¹⁶ Algebraan perustuvia edustavat vektoriavaruusmalli sekä erilaiset neuroverkkoihin perustuvat lähestymistavat, todennäköisyysteoriaan perustuviin kuuluvat muun muassa probabilistinen täsmäytys sekä kielimallit.

Yleisesti ottaen hakutuloksiin vaikuttavat joka tapauksessa useat muuttujat. Lisäksi on olemassa niin sanottuja metahakukoneita, jotka yhdistelevät useiden hakukoneiden tarjoamia hakutuloksia. Hakualgoritmien toiminta on tyypillisesti salattu käyttäjiltä ainakin osittain. Ilmeisesti ainakin kaupalliset hakukoneyhtiöt pyrkivät vaalimaan liikesalaisuuksia kilpailijoiltaan ja pyrkivät estämään pahantahtoisen *black hat* -hakukoneoptimoinnin. Hakualgoritmien salaaminen on kuitenkin herättänyt kysymyksiä siitä, missä määrin ja millä perusteella hakukoneyhtiöt muokkaavat käyttäjälle tarjottavia hakutuloksia.

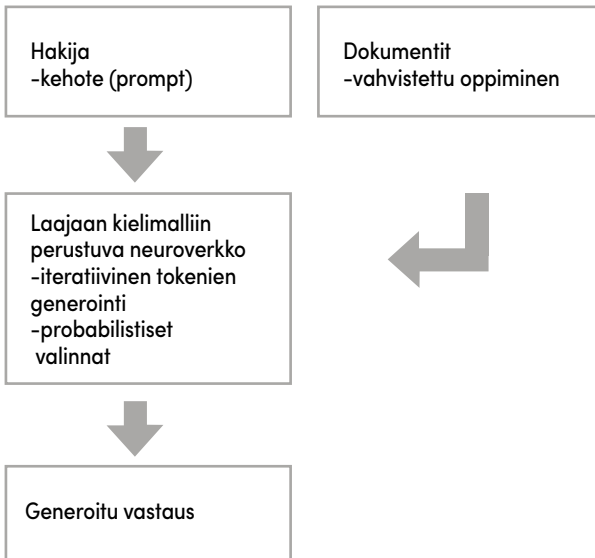
Informaation generointi

Tässä kirjassa monelta suunnalta esitellyn informaatiotutkimuksen tavoitteeksi on aikoinaan lausuttu, että sen *"tavoitteena on kehittää käsitteitä, menetelmiä ja järjestelmiä, joiden avulla tieto saadaan vaittomasti sitä tarvitsevan ulottuville helposti omaksuttavassa muodossa."*¹⁷ Olemassa olevat hakukoneet toteuttavat tätä tavoitetta kuitenkin vain osittain. Esimerkiksi ne tarjoavat tiedon lähinnä listana viitteitä *dokumentteihin*. Tällöin hakutulosten avulla hakijan on tyypillisesti *itse koostettava* ja jalostettava näin saatu tieto haluttuun muotoon.

Vuoden 2023 alkupuolella suuren yleisön tietoisuuteen nousivat nopeasti laajoihin generatiivisiin kielimalleihin (Large Language Models, LLM) pohjautuvat sovellukset, eturintamassa GPT-malliin (Generative Pre-trained Transformer) perustuva ChatGPT. Siinä,

¹⁶ Croft ym. 2010.

¹⁷ Ks. Mäkinen 2010.



Kuvio 4. Informaatiota muodostavan hakujärjestelmän yleinen malli.

missä perinteiset hakukoneet noutavat tietokantaan tallennettuja dokumentteja sellaisenaan, informaatiota generoivat järjestelmät muodostavat tekstiä dynaamisesti vastauksena hakijan antamaan kehotetekstiin hyödyntäen algoritmisesti lukuisia tekstilähteitä.

Kuviossa 4 pelkistetyt esitetyn kielimallisovelluksen syöte ja tuloste lähestyvät kuitenkin informaatiotutkimuksen idealisoitua tavoitetta, koska *muodostuva informaatio* ainakin periaatteessa saadaan sitä tarvitsevan ulottuville jopa *valmiissa muodossa*, esimerkiksi tekstinä (proosa) tai ohjelmakoodina. Kielimallisovelluksen anatomia onkin karkealla tasolla jossain määrin samantyyppinen hakukoneen kanssa, sillä siinäkin käyttäjän antamaa tiedontarpeen ilmaisua vertaillaan dokumenteista luotua tietosisältöä vasten. Kielimallien yhteydessä tiedontarve ilmaistaan hakukonekyselyn sijaan ennemminkin luonnollisen kielen *kehotteena (prompt)* ja tulostetta kutsutaan hakutuloksen sijaan toisinaan tuotokseksi, vastaukseksi tai generoiduksi tekstiksi.

Kyselyiden prosessointia kutsuttiin hakukoneiden yhteydessä *täsmäytykseksi*. Vastaavasti esimerkiksi GPT-kielimallisovelluksen toiminta perustuu niin sanottuun transformer-arkkitehtuuriin.¹⁸ Arkkitehtuuri on syväoppimismalli, joka käsittelee luonnollista kieltä ja käyttää kerrostettua neuroverkkoarkkitehtuuria. Transformer-arkkitehtuurissa on kaksi pääosaa: koodaaja (*encoder*) ja purkaja (*decoder*). Koodaaja lukee syötetekstin ja muuntaa sen tietokoneen ymmärtämään taulukkomuotoon *neuroverkoksi*. Purkaja avaa tämän tiedon ja muuttaa sen takaisin luonnolliseksi kieleksi. Oleellinen osa mallia on pyrkiä tunnistamaan kehoitteista ja lähdeteksteistä keskeiset termit.¹⁹

Edellä mainittujen teknisten piirteiden lisäksi eräs eroavaisuus informaatiota hakevien hakukoneiden ja informaatiota generoivien kielimallisovellusten välillä liittyy informaatiovuorovaikutuksen luonteeseen. Hakukoneen käyttäjä on vuorovaikutuksessa hakujärjestelmän ja sen tarjoamien dokumenttien kanssa ja suorittaa yleensä useita hakuja tehtävän suorittamiseksi osana *tiedon hankintaprosessia*. Hakukone ei kuitenkaan tyypillisesti muista tai varsinaisesti ota huomioon käyttäjän aiempia hakuja. Hakija tyypillisesti etsii lisää dokumentteja muokkaamalla kyselyään ja tutkimalla näin saatua uutta, edellisistä hauista riippumatonta tuloslistaa.

Sen sijaan informaatiota generoivissa kielimallisovelluksissa vuorovaikutus muistuttaa enemmän keskustelua, jossa sovellus ottaa huomioon prosessoinnissaan myös saman istunnon aikaiset aiemmat kehoitteet. Tästä juontuu nimi ChatGPT: ohjelma simuloi chat-keskustelua, jonka suuntaa käyttäjät voivat *ohjata* kehoittein. Käyttäjä voi syöttää kehoitteita kokeilevassakin hengessä, mikä on hyvä alku, mutta parhaaseen lopputulokseen pääsemiseksi tarvitaan suunnitelmallisempaa ja systemaattisempaa, jopa ammattimaisempaa lähestymistapaa. Hyvällä *kehotesuunnittelulla* (*prompt engineering*) voidaankin tehokkaasti ja nopeasti suorittaa tarkasti

¹⁸ Ks. Wolfram 2023.

¹⁹ Vashvani ym. 2017.

määriteltävissä olevia tehtäviä tai ratkaista erittäin monimutkaisia haasteita vaiheittain *ajatusketjujen* avulla.²⁰ Ajatusketjuissa käydään dialogia tekoälysovelluksen kanssa. Tehtävän suoritusta ohjataan ja mallia koulutetaan suunnitelmallisesti ja intellektuaalisesti haluttuun suuntaan. Kehotesuunnittelu on siis iteratiivista informaatiovuorovaikutusta käyttäjän, tekoälyn ja opetusdatassa olevan laajan aineiston välillä.

Generoitavan lopputuloksen laatu riippuu käyttäjän kyvystä arvioida saatua vastausta ja sen pohjalta laatia uusia täsmällisiä ja soveltuvia kehoitteita. Onnistunut kehotesuunnittelu ja tulosten hyödyntäminen edellyttävät näin ollen syvällistä osaamista tehtävän alalta. Tämä koskee erityisesti asiantuntemusta vaativien tehtävien suorittamista.²¹ Kielimallin perusteella generoidun informaation tarkkaa alkuperää ei voida jäljittää, jolloin käyttäjän on kyettävä arvioimaan vastaustekstien oikeellisuutta, ajantasaisuutta ja relevanssia puhtaasti omista tiedollisista lähtökohdistaan. Kielimalliin perustuvaa virheellistä, mutta uskottavaa generointia kutsutaan toisinaan tekoälyn *hallusinoinniksi*.

Kielimalleja käyttävien sovellusten arkkitehtuurissa generointia pyritään kohdentamaan ja hallusinointia suitsimaan erityisesti käyttäen *tiedonhakuavusteista generointia* (Retrieval Augmented Generation, RAG).²² Siinä kielimallisovelluksen osa, hakulohko (Retriever), hakee generoinnin tueksi relevanttia, kohdennettua ja ajantasaista tietoa määrätystä ulkoisesta tietokannasta, kuten dokumenttivarastosta, tietokannasta tai internetistä. Haun tulos yhdistetään automaattisesti muuhun sovellukselle tallennettuun ja syötettyyn dataan, kuten käyttäjän kehoitteisiin. Tämän kokonaisuuden perusteella sovellus generoi tuotoksen suuren kielimallin avulla. Hakulohko on olennainen osa useiden kielimalleja käyttävien sovellusten arkkitehtuuria, ja sen toiminta on tässä luvussa kuvatus hakukoneen kaltainen. Usein hakulohko käyttääkin jonkin olemassa olevan ha-

20 Wei ym. 2022.

21 Esim. Sinha ym. 2023.

22 Esim. Lewis ym. 2020.

kukoneen tarjoamaa *rajapintaa*, johon sovellus muotoilee esimerkiksi generoimalla kyselyn tiedonhakua varten.

Kielimallit ovat rakenteeltaan monimutkaisia verkostoja ja tarvitsevat toimiakseen – edes tyydyttävästi – valtavan määrän tekstimassaa opetusdataksi. Aineisto voi kattaa erittäin laajan otoksen saatavilla olevasta julkaistusta tekstidatasta. Hakukone sen sijaan toimii aiemmin tässä luvussa kuvatuilla, melko yksinkertaisilla toiminnoilla hyvinkin rajatussa ympäristössä, esimerkiksi työpaikan intranetissä. Kielimallien monimutkaisuuden sekä massiivisen datamäärän takia tiedon hakijan on hyvin vaikea ymmärtää kielimallin toimintaa ja saatujen tulosten alkuperää, vaikka hän olisi itse järjestelmän suunnitellut insinööri. Erityisesti aiemmin kuvatussa täysmäyttävässä hakukoneessa osaavalla hakijalla on kontrolli hakutuloksiin. Hyvin määritelty hakualgoritmi indeksointiprosesseineen tekee hakuprosessista läpinäkyvän ja sen vuoksi käyttökelpoisen tarkkaa kontrollia vaativiin toimintoihin, kuten dokumentoituun aineiston keräämiseen tutkimuksessa tai vaikkapa kattavaan patenttien hakuun. Toisaalta generoinnin tiedonhakupohjainen laajentaminen parantaa tarvittaessa generointiprosessin läpinäkyvyyttä ja tarjoaa tiedolle mahdollisia lähteitä.

Tämän luvun lähtökohtana esiteltiin tiedonhaun malli, joka koostui neljästä komponentista: hakijasta tiedontarpeen ilmaisuiheen (kyselyt), haettavista dokumenteista tai informaatioyksiköistä, jotka hakujärjestelmä prosessoi haettaviksi ilmaisuiksi, hakualgoritmista, joka vertailee tiedontarpeen ja dokumenttien ilmaisuja, sekä hakutuloksesta, jonka hakualgoritmi muodostaa (esimerkiksi linkkilista). Perinteisten dokumenttienhakujärjestelmien rinnalla toimimme esiin kielimalleihin ja tekoälyyn perustuvat, suuria neuroverkkoja hyödyntävät, informaatiota muodostavat eli generoivat järjestelmät. Erityyppisten järjestelmien kehittäjille, tiedon hakijoille, dokumenttien kirjoittajille ja tiedonhaun tutkijoille on tärkeää, kuinka hyvin tiedon haku tai muodostaminen onnistuu ja mitkä tekijät tähän vaikuttavat. Voidaan esimerkiksi kysyä, kuinka tiedontarpeen muotoilu (kyselyin tai kehoitein) ja haussa käytettävä algoritmi vaikuttavat

siihen, kuinka relevanttia informaatiota hakujärjestelmä palauttaa. Tällöin kohtaamme haastavan kysymyksen relevanssista.

Tiedonhaun tutkimus kohtaa alati uusia haasteita. Moninaista todellisuutta erilaisista näkökulmista redusoivien mallien käyttöön tutkimuksessa sisältyy tutkijan kannalta paitsi kiinnostavia mahdollisuuksia myös rajoitteita ja metodologisia kehittämistarpeita.

Lähteet

- Alkula, Riitta (2000) *Merkkijonoista suomen kielen sanoiksi*. Väitöskirja. Tampere: Tampereen yliopisto.
- Azzopardi, Leif (2007) Position paper: Towards evaluating the user experience of interactive information access systems. *SIGIR'07 web information-seeking and interaction workshop*.
- Belkin, Nicholas J. (1980) Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information and library science* 5, 133–143.
- Brin, Sergey & Lawrence Page (1998) The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30:1–7, 107–117.
- Croft, W. Bruce & Donald Metzler & Trevor Strohman (2010) *Search engines: Information retrieval in practice*. Boston: Addison Wesley.
- Duk, Sanja & Dunja Bjelobrk & Mia Čarapina (2013) SEO in e-commerce: balancing between white and black hat methods. *MIPRO 2013*, May 20–24, 2013, Opatija, Croatia, 390–395.
- Ingwersen, Peter & Järvelin Kalervo (2005) *The Turn: Integration of information seeking and retrieval in context (The information retrieval series)*. Dordrecht: Springer.
- Jansen, Bernard J. & Amanda Spink & Tefko Saracevic (2000) Real life, real users, and real needs: A study and analysis of user queries on the web. *Information processing & management* 36:2, 207–227.
- Järvelin, Kalervo & Pertti Vakkari & Paavo Arvola & Feza Baskaya & Anni Järvelin & Jaana Kekäläinen & Heikki Keskustalo & Sanna Kumpulainen & Miamaria Saastamoinen & Reijo Savolainen & Eero Sormunen (2015) Task-based information interaction evaluation: The viewpoint of program theory. *ACM transactions on information systems* 33:1, nro 3.
- Kuhlthau, Carol C. (1991) Inside the search process. *Journal of the American society for information science* 42:5, 361–371.
- Lewis, Patrick & Ethan Perez & Aleksandra Piktus & Fabio Petroni & Vladimir Karpukhin & Naman Goyal & Heinrich Küttler & Mike Lewis & Wentaui Yih & Tim Rocktäschel & Sebastian Riedel & Douwe Kiela (2020) Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in neural information processing systems* 33, 9459–9474.

- Mäkinen, Ilkka (2010) Informaatiotutkimuksen tie. Teoksessa Sami Serola (toim.) *Ote informaatiosta*. Helsinki: BTJ Kustannus, 11–74.
- Ruthven, Ian (2008) Interactive information retrieval. *Annual review of information science and technology* 42, 43–91.
- Sinha, Ranvir K. & Asitava Deb Roy & Nikhil Kumar & Himel Mondal (2023) Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus* 15:2.
- Spärck Jones, Karen (1972) A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation* 28:1, 11–21.
- Vakkari, Pertti (2000) Cognition and changes of search terms and tactics during task performance: a longitudinal study. *Proceedings of the RIAO 2000 conference*, Paris: C.J.D., 894–907.
- Vakkari, Pertti (2002) Subject knowledge, source of terms, and term selection in query expansion: an analytical study. *Proceedings of the 24th European colloquium on IR research (ECIR'02)*, LNCS 2291, 110–123.
- Vaswani Ashish & Noam Shazeer & Niki Parmar & Jakob Uszkoreit & Llion Jones & Aidan N. Gomez & Lukasz Kaiser & Illia Polosukhin (2017) Attention is all you need. arXiv preprint. <https://arxiv.org/pdf/1706.03762.pdf>
- Wei, Jason & Xuezhi Wang & Dale Schuurmans & Maarten Bosma & Brian Ichter & Fei Xia & Ed Chi & Quoc Le & Denny Zhou (2022) Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35, 24824–24837.
- Wolfram, Stephen (2023) *What is ChatGPT doing ... and why does it work?* Champaign: Wolfram Media.

3.2 Tiedonhaun tutkimus

Heikki Keskustalo, Paavo Arvola ja Sanna Kumpulainen

- Tiedonhaun tutkimuksen keskeisenä tavoitteena on ymmärtää tiedonhakua ilmiönä.
- Tyypillisiä tutkimuskohteita ovat hakija, dokumentit, hakualgoritmit ja haettu tai muodostettu informaatio.
- Tiedonhaun tutkimustyytit eroavat toisistaan keskeisesti siinä, millaisena *hakija* tutkimuksissa näyttäytyy. Tutkimuksessa läsnä oleva hakija voi olla autenttinen toimija tai koehenkilö suorittamassa annettua tehtävää – tai läsnä vain välillisesti.
- Hakijan läsnäolon tapa tutkimuksessa vaikuttaa siihen, millaisia kysymyksiä voidaan tutkia.
- Tiedon kenttä muuntuu alati tarjoten tutkimukselle jatkuvasti uusia kysymyksiä.

Tiedonhaun tutkimuksessa tarkastellaan, kuinka erilaiset tietojärjestelmät tuottavat relevanttia informaatiota hakijalle. Oppialana tiedonhaku voidaan sijoittaa informaatiotutkimuksen ja tietojenkäsittelytieteen rajamaastoon. Alan pioneerien Saltonin ja McGillin¹ mukaan tiedonhaussa ollaan kiinnostuneita *informaatioyksiköiden* (*information items*, ks. luku 2.3) esittämisestä, tallennuksesta, organisoinnista ja hakemisesta. Perinteisen dokumenttien haun rinnalle kasvavan mielenkiinnon kohteiksi ovat viime aikoina nousseet informaatiota *muodostavat* eli generoivat järjestelmät, kuten ChatGPT.² Haettaessa “valmiita” informaatioyksiköitä tai *muodostetta*ssa informaatiota dynaamisemmin voimme kummassakin tapauk-

¹ Salton & McGill 1987.

² Ks. esim. Wolfram 2023.

sessä puhua tiedonhausta. Tiedonhaun käyttöalat, tekniikat ja käytötavat muuntuvat alati ja esimerkiksi web-sovelluksia käytettäessä tiedonhaku sulautuu usein huomaamattomaksi osaksi erilaisia arjen palveluita.

Vaikka sana *tiedonhaku* viittaa vain tiedon hakemiseen, tutkimusalan nimityksenä termi kattaa laaja-alaisen joukon ilmiöitä. Tiedonhakua tutkittaessa hakijan tyypillisesti ajatellaan olevan kiinnostunut *relevantin* informaation löytämisestä tai muodostamisesta – ei dokumenttien hakemisesta sinällään. Tällöin dokumenttien ajatellaan *sisältävän* tulkittavissa olevaa informaatiota.

Tiedonhaun tutkimusasetelmat

Tiedonhaun tutkimuksen yleisenä tavoitteena on ymmärtää tiedonhakua ilmiönä ja kehittää parempia hakujärjestelmiä ja -menetelmiä. Tämän kirjan luvussa 3.1 esitetään tiedonhaun yleinen malli, jonka pääelementit ovat hakija, dokumentit, hakualgoritmi sekä haun lopputulos eli haettu informaatio. Näiden elementtien avulla perinteinen tiedonhakuprosessi voidaan pelkistää seuraavasti: (i) hakija ilmaisee tiedontarpeensa hakujärjestelmälle käyttäen apuna kyselyä, (ii) järjestelmä antaa tarjolle dokumentteja haettaviksi sopivia ”tietorakenteita”, kuten hakemistoa, hyödyntäen, (iii) algoritmi vertailee tiedontarpeen ja dokumenttien ilmaisuja ja (iv) hakutulokset esitetään hakijalle, esimerkiksi linkkilistana.

Käytännössä tiedonhakijat ja hakutilanteet voivat olla hyvin monentyyppisiä. Tiedonhaun *tutkimus* voidaan usein nähdä hakutilanteiden *simulaationa*, jossa tosiasiallista hakutilannetta jäljitellään ja tilanteen valikoituja piirteitä tarkastellaan lähemmin. Yksittäisessä tutkimuksessa käytetty hakijan ja hakuprosessin malli voi tällöin olla karkean yksinkertainen tai huomattavan hienojakoinen ja monimutkainen. Pyrimme seuraavassa valaisemaan asiaa luomalla kokonaiskuva erityyppisistä tutkimuksista käyttäen esimerkkejä. Haluamme korostaa *erilaisia* mahdollisuuksia huomioida hakujärjestelmän käyttäjä, kun tiedonhakua tutkitaan.

Yksittäisessä tutkimuksessa voidaan aina huomioida vain rajattu joukko hakijan piirteitä. Tiedonhaun tutkimusasetelmat voidaan karkeasti jakaa kolmeen tyyppiin, jotka kuvastavat eri tapoja suhtautua hakijaan. Suhtautumistavat voidaan otsikoida seuraavasti: (1) *autenttisten hakutilanteiden tutkiminen*, (2) *koehenkilöt suorittamassa heille annettuja tehtäviä* ja (3) *testikokoelmien hyödyntäminen*. Tyyppien 1 ja 2 tutkimuksissa tarvitaan tutkimukseen osallistuvia ihmisiä, joiden toimintaa tarkastellaan. Kolmanteen tyyppiin kuuluvia tutkimuksia voidaan sen sijaan tehdä ilman testihenkilöitä, hyödyntämällä *testikokoelmaa*.

Autenttisten hakutilanteiden tutkiminen

Autenttinen tiedonhakutilanne tarkoittaa todellisen maailman tiedonhakua, jossa ihmiset tekevät tiedonhakuja luonnollisissa ympäristöissä, kuten työpaikoilla, liittyen oikeisiin tehtäviin, ongelmiin ja tilanteisiin. Tällainen tutkimus on harvinaista, sillä sen toteuttaminen on vaivalloista ja vaatii pääsyn erilaisiin tilanteisiin sekä runsaasti aikaa. Todellisen maailman asetelmilla voidaan kuitenkin löytää ja saada esille sellaista, mikä jäisi tutkimusasetelmien ulkopuolelle, jos asetelmat olisivat liian rajattuja. Tällaisia asioita voivat olla esimerkiksi tiedostamattomat tiedonhakutavat ja oikopolut, joista on tullut rutiineja työssä eikä niihin välttämättä kiinnitetä tietoisesti huomiota.³ Tiedonhaun tutkimus lähestyy tällöin tiedonhankinnan tutkimusta ja tietokäytäntöjen tutkimusta, sillä huomioon otetaan tietojärjestelmien lisäksi myös muita tiedonlähteitä. Näitä voivat olla esimerkiksi painetut lähteet, omat muistiinpanot ja muut ihmiset, joilta kysytään neuvoja.

Usein autenttisia tiedonhakutilanteita tutkitaan työpaikoilla ja -tilanteissa. Informaatio näyttäytyy näissä varsin monimuotoisena. Työn muutos ja digitalisaatio ovat muuttaneet informaatioympäristöjä erilaisilla työpaikoilla. Hyvän informaatiotutkimuksen näkökul-

3 Kumpulainen 2017; Kumpulainen 2013.

masta tehdyn jäsenyyksen informaation roolista työpaikoilla tarjoaa Byströmin ja kumppanien teos⁴, jossa aiheen tutkimusta käsitellään monipuolisesti. Tutkimuksia on tehty eri aloilla, muun muassa sosiaalityön puitteissa,⁵ kaupungin hallinnossa⁶ sekä historian tutkijoiden työtehtävien parissa.⁷ Kaikissa näissä eri alojen vaatimukset tietojärjestelmille ja niiden sisältämälle informaatiolle näyttäytyvät käyttötilanteiden kautta jokseenkin erilaisina. *Työn tavoitteet* määrittävät informaation hyödyllisyyttä. Tarkastelemme teemaa lisää luvussa 2.3, joka käsittelee informaatiovuorovaikutusta.

Autenttisten hakutilanteiden tutkimuksissa voidaan tarkastella myös vuorovaikutus- ja tiedonhakulokeja muun tutkimusaineiston (havainnointi, haastattelut ja lokit) tukena⁸ tai tarkastella lokeja itessään,⁹ jolloin tutkimuksen kohteena on usein suppeampi haku-tehtävä kuin edellä kuvatuissa työpaikoilla tehdyissä tutkimuksissa. Tällöin tarkastelutasona on usein työtehtävän kokonaisuus tai sen osa.

Koehenkilöt suorittamassa annettuja tehtäviä

Autenttisten hakutilanteiden tutkimusta voidaan pitää tietyllä tapaa ihanteellisena. Parhaimmillaan asetelmasta puuttuu keinotekoisuus ja tutkija voi itse valita, millaisiin kysymyksiin hän keskittyy. Kaikkia tutkijaa kiinnostavia kysymyksiä ei kuitenkaan välttämättä voi tutkia autenttisissa tilanteissa. Sopivia tilanteita ei välttämättä ole tutkijalle saatavilla. Tällaisissa tilanteissa voidaan hyödyntää *simuloituja tehtäviä*.¹⁰ Niissä koehenkilöt suorittavat autenttisten tehtävien sijasta *tutkijan antamia* tehtäviä. Tehtävänanto koostuu *taustatarinasta* sekä yksityiskohtaisemmista toimintaohjeista (kuvio 1).

4 Byström ym. 2019.

5 Huuskonen & Vakkari 2010.

6 Byström & Järvelin 1995; Saastamoinen & Kumpulainen 2014.

7 Korkeamäki & Kumpulainen 2019; Kumpulainen ym. 2020.

8 Ks. esim. Kumpulainen & Järvelin 2010; Saastamoinen & Järvelin 2017.

9 Odijk ym. 2015; White & Huang 2010.

10 Borlund 2016.

Taustatarina

- Tehtävä: kuvittele, että olet kirjoittamassa artikkelia, joka käsittelee Suomen historiaa tai maailmanhistoriaa 1800-luvun lopulla tai 1900-luvun alun näkökulmasta.
- Ohje relevanssiarvion tekemiseen on kuvattu alla.

Hakutuloksen evaluointi

- Relevanssitaso 3: Näkemäsi lehtileike sisältää runsaasti informaatiota hakemastasi aiheesta ("Bobrikoffin murha") ja lehtileike auttaa hyvin tehtävän suorittamisessa.
- Relevanssitaso 2: Lehtileike koskettaa hakemaasi aihetta ja auttaa jossain määrin tehtävän suorittamisessa.
- Relevanssitaso 1: Lehtileike liittyy aiheeseen, mutta se ei välttämättä auta suoraan tehtävän suorittamisessa.

Kuvio 1. Simuloidun tehtävän kuvaus (muokattu Kettusen ja muiden pohjalta).

Kuvio 1 havainnollistaa *simuloitua* hakutehtävää.¹¹ Koehenkilön tehtävänä on hakea hänelle annettua hakukonetta käyttäen digitoituja historiallisia sanomalehtileikkeitä ja *arvioida lehtileikkeiden hyödyllisyyttä taustatarinan valossa*. Kyseisessä tutkimuksessa paperille painetut historialliset sanomalehtiartikkelit näytettiin hakijalle digitaalisessa muodossa. Painettu teksti on kuitenkin jouduttu tunnistamaan optista merkintunnistusta käyttäen. Muodostettuun digitaaliseen tekstiin syntyy samalla tunnistusvirheitä. Tekstissä esiintyvät virheelliset kirjainmerkit haittaavat tiedonhakua sekä dokumenttien lukemista. Tutkimuksessa vertailtiin kahden eri merkintunnistumenetelmän (*optical character recognition*, OCR) vaikutuksia hakuun. Vertailussa olivat perusmenetelmä sekä vähemmän merkkivirheitä muodostava kehittyneempi versio. Koehenkilön käyttämä hakukone kykeni esittämään dokumentin kahtena vaihtoehtoisena versiona (joko perus- tai kehittyneempi OCR) ja keräämään talteen hakijan antaman palautteen dokumentin hyödyllisyydestä (hakija klikkasi dokumentille "laatuarvosanan" 0, 1, 2 tai 3). Tämä simuloitu

11 Kettunen ym. 2022.

Hakuaiheet	Dokumenttikokoelma	Relevanssikorpus
Topic-id: #15 Aihe: Viiriäiset Suomessa Kuvaus: Relevantti (Rel=1) dokumentti sisältää tietoja viiriäistä koskevista havainnoista Suomessa ... Epärelevantti dokumentti (Rel=0) sisältää ... Topic-id ...	Docid: #300700 Otsikko: "Oletko nähnyt viiriäistä" Dokumentin teksti (5673): " ... kokkimme näki omin silmin viiriäisen käydessään ..." Docid: 300701 ...	Topic-id Docid Rel ... #15 #300601 0 #15 #300700 1 ... #16 #101315 0 ...

Kuvio 2. Testikokoelman komponentit: hakuaiheet, dokumentit ja relevanssiarviot.

asetelma teki mahdolliseksi tutkia, kuinka OCR-virheiden väheneminen vaikuttaa dokumenttien koettuun hyödyllisyyteen. Ilman simuloitua asetelmaa tällaista kysymystä olisi vaikea tutkia.¹²

Testikokoelmatutkimus

Testikokoelmatutkimus havainnollistaa kolmatta tapaa suhtautua tiedonhakijaan tutkimuksessa. Tätä jo 1960-luvulla niin sanotuissa Cranfield-kokeissa käytettyä mallia on hyödynnetty runsaasti tiedonhaun tutkimuksissa.¹³ Testikokoelmatutkimuksessa koehenkilöitä ei välttämättä käytetä lainkaan. Tiedonhaun “testiajoja” voidaan suorittaa esimerkiksi omalla tietokoneella. *Testikokoelma* koostuu kolmesta rakenneosasta, jotka ovat (i) hakuaiheet (*topics*), (ii) dokumentit (*documents*) ja (iii) relevanssikorpus (*relevance corpus*). *Hakuaiheet* määrittelevät sen, mitä hakujärjestelmästä tarkkaan

12 Kettunen ym. 2022.

13 Cleverdon 1991; Voorhees 2007.

2 kyselyä (topic id #15)	Hakutulos 1 (q1)	Hakutulos 2 (q2)
- q1: viiriäiset Suomessa - q2: viiriäiset	- 300700 Rel=1 - 300601 Rel=0	- 300601 Rel=0 - 300700 Rel=1

Kuvio 3. Esimerkki testikokoelman hyödyntämisestä. Kahden kyselyn (q1 ja q2) tuottamia hakutuloksia voidaan vertailla hyödyntäen relevanssikorpuksen relevanssitietoja.

ottaen pyritään hakemaan. Hakuaiheita on tyypillisessä testikokoelmassa vähintään kymmeniä – tai jopa satoja. *Dokumentit* eli dokumenttitietokanta muodostavat hakujen kohteen. Tietokanta sisältää yleensä laajan joukon dokumentteja, ja vain hyvin pieni osuus dokumenteista on relevantteja jonkin yksittäisen hakuaiheen näkökulmasta. Kolmas rakenneosa, *relevanssikorpus*, sisältää relevanssiarviot eli tiedon siitä, mitkä dokumentit ovat tosiasiaa relevantteja kullekin hakuaiheelle (kuvio 2).

Kuviossa 2 on havainnollistettu testikokoelman komponentteja. Vasemmalla näkyvä *hakuaihe* (”Viiriäiset Suomessa”) määrittelee haun aiheen ja kuvaa, millainen dokumentin tulee olla, jotta se olisi relevantti. Kuviossa keskellä on havainnollistettu esimerkinomaisesti kokoelman yksittäistä dokumenttia. Hakuaihe ja dokumentti on merkitty yksilöivillä koodeilla (#15 ja #300700). Oikealla oleva *relevanssikorpus* kertoo, mitkä dokumentit ovat relevantteja kullekin aiheelle. Esimerkin ylin numerorivi kertoo, että *viiriäisaiheelle* (#15) dokumentti numero 300601 ei ole relevantti, sillä sen relevanssitaso on 0. Toinen rivi kertoo vastaavasti, että dokumentti numero 300700 on aiheelle relevantti, sillä sen relevanssitaso on 1. Testikokoelma rakennetaan käytännössä valitsemalla kohdedokumentit ja laatimalla sopivat hakuaiheet sekä relevanssiarviot. Relevanssiarviot muodostetaan niin sanotulla tuomarimenetelmällä, jolloin ihmiset lukevat dokumentteja ja arvioivat, ovatko ne relevantteja suhteessa valittuun aiheeseen. Testikokoelman valmistuttua sitä voidaan hyödyntää lukuisin tavoin tutkimuksessa. Yleisperiaatteena on muo-

dostaa kullekin hakuaiheelle kyselyitä ja hakea niiden avulla dokumentteja. Olennaista on oivaltaa, että hakutulosten *onnistuneisuutta* voidaan aina mitata *relevanssikorpuksen* ansiosta (kuvio 3).

Kuvio 3 havainnollistaa hakutulosten tarkastelua. Antoiko kysely q1 vai q2 paremman hakutuloksen? Kysely q1 palautti dokumentit #300700 ja #300601 tässä järjestyksessä ja (lyhyempi) kysely q2 palautti samat dokumentit, mutta päinvastaisessa järjestyksessä. Hakutuloksiin on kuviossa liitetty automaattisesti relevanssikorpuksen sisältämiä tietoja, jolloin tulosten paremmuutta voidaan verrata. Havaitsemme, että kyselyversio q1 palautti relevantin dokumentin ykkössijalla ja epärelevantin kakkossijalla. Intuitiivisesti arvioiden kysely q1 antoi hieman paremman tuloksen kuin q2. Tulosten täsmälliseen vertailuun on olemassa lukuisia numeerisia mittareita, jotka mallintavat esimerkiksi hakijan erilaisia preferenssejä. Yksi esimerkki tällaisesta mittarista on hakijan kärsivällisyys tutkia hakutulosta. Kuvio 3 havainnollistaa, kuinka testikokoelman avulla voitaisiin verrata esimerkiksi samasta aiheesta tehtyjen eri kyselyversioiden vaikutusta hakutulokseen.

Testikokoelmia hyödynnettäessä tehdään tyypillisesti tiedonhakuja kullakin (vähintään kahdella) menetelmällä, joita halutaan verrata, käyttäen useita hakuaiheita, pyrkimyksenä muodostaa yleistettäviä johtopäätöksiä. Tutkijaa kiinnostava kysymys voisi olla esimerkiksi, millaisia hakutuloksia lyhyet kyselyt tuottavat verrattuina pitkiin kyselyihin. Tällaista tutkimusasetelmaa on myös pelillistetty siten, että pelaajien syöttämät erilaiset kyselyt kilpailevat siitä, mitkä niistä tuottavat parhaita tuloksia.¹⁴

Kyselyjen ohella voidaan kokeellisesti tutkia myös esimerkiksi sitä, kuinka hyvin informaatiota *muodostavat* kielimallit, kuten ChatGPT, suoriutuvat inhimillisessä päätöksenteossa, vaikkapa ratkaistessaan oikeustapauksia.¹⁵ Erilaisten kehoitteiden tai *kehotestategioiden* kuten ajatusketjujen *lopputulosta* päätöksentekoprosessissa

14 Arvola & Alamettälä 2022a; 2022b; Sormunen ym. 2002; Sormunen ym. 1998.

15 Esim. Yu ym. 2023.

voidaan yksinkertaisimmillaan evaluoida *tuomioistuinten tekemiä* langettavia tai vapauttavia *päätöksiä vasten*. Saatua tulosta analysoidaan määrällisesti eli kuinka usein spesifi kehotestrategia johtaa oikeaan ratkaisuun. Tämän kaltainen tutkimus on luonteeltaan *kysymys-vastaus*-tyyppisen (*Question Answering, QA*) tiedonhaun evaluointia, jossa hakukoneelle esitetään kysymys luonnollisella kielellä ja vastausten laatua arvioidaan verrattuna ihmisten laatimiin vastauksiin tai ratkaisuihin.¹⁶

Testikokoelmaa voidaan hyödyntää paitsi *kyselyiden* ja *kehotteiden* keskinäiseen vertailuun (tiedontarpeen ilmaisutapojen menestyksellisyys) myös *dokumenttien* ilmaisutapojen vertailuun (esimerkiksi kannattaisiko hakujärjestelmässä asettaa sanat haettaviksi taipuneissa muodoissaan tai perusmuotoisina) sekä *hakualgoritmien* vertailuun. Siinä, missä arjen tiedonhaussa tiedontarpeen ilmaisu, dokumenttien tietosisällön ilmaisu ja hakualgoritmi muodostavat erottamattoman kokonaisuuden, testikokoelman avulla voidaan tutkia kokonaisuuden yksittäisten osien vaikutusta haun tuloksiin. Perusidea on tällöin varioida yksittäistä asetelman komponenttia kerrallaan (vaikkapa *kyselyiden* tyyppiä) pitäen samalla asetelman muut pääkomponentit ennallaan (esimerkiksi indeksin tyyppi ja kyselyitä dokumentteihin vertaileva hakualgoritmi).

Testikokoelmaa hyödynnettäessä näkemys *hakijasta* pelkistyy voimakkaasti. Hakutilanteiden vertailu perustuu usein *yksittäisten* kyselyiden käyttöön hakuaiheille, jotka ovat täsmällisesti määriteltäviä, evaluoinnissa tarkastellaan haun lopputuloksia, relevanssi perustuu hakuaiheen ja dokumentin *aiheisältöjen* vastaavuuteen ja dokumentin relevanssi arvioidaan riippumatta muista samanaikaisesti haetuista dokumenteista. Todellisuudessa hakuprosessi on kuitenkin yleensä *olennaisesti vuorovaikutteinen*. Hakija voi turvautua useampiin kyselyihin, hakuaihe voi olla epäselvä ja hakijan relevanssikriteerit voivat olla subjektiivisia ja muuttuvia. Lisäksi hakija voi arvioida haettuja dokumentteja kokonaisuutena, ja muutkin seikat

16 Webber 2010; Voorhees & Tice 2000.

kuin hakutulosten relevanssi voivat olla olennaisia – esimerkiksi hakemisen vaivattomuus ja nopeus.

Testikokoelmia voidaan hyödyntää varsin monentyyppisen hakijoiden toiminnan tutkimiseen.¹⁷ Esimerkkinä *monen kyselyn hakuistuntojen* simuloinnissa kuhunkin hakuaiheeseen liitetään yksittäisen kyselyn sijasta erilaisten kyselyiden sarja q1, q2, q3 ja niin edelleen. Tällöin testikokoelmaa käyttäen voidaan simuloida tilannetta, jossa hakija suorittaa hakuaiheelle useita kysely-yrityksiä selaillen hakutulosta kunkin kyselyn jälkeen. Tällaisen aselman erikoistapauksena voidaan pitää *relevanssipalautteen simulointia*, jolloin haetuista dokumenteista eristetään automaattisesti sanoja alkuperäiseen kyselyyn lisättäviksi. Testikokoelman käyttöalaa voidaan tähän tapaan pyrkiä laajentamaan, jolloin kokoelmaa hyödynnetään esimerkiksi sellaisen toiminnan jäljittelyyn, jossa hakija tarkastelee hakutulosta ja parantelee kyselyään lisäämällä siihen hyviä hakusanoja.

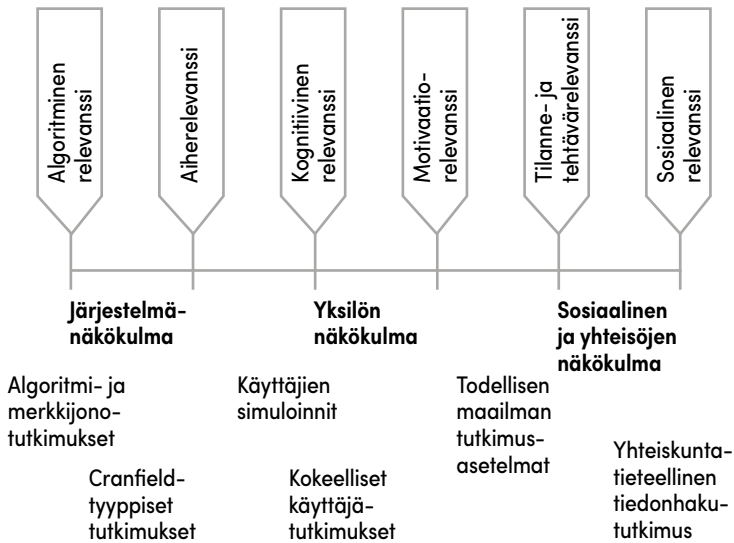
Testikokoelmien käytön etuna on, että niiden avulla voidaan systemaattisesti tarkastella hakujärjestelmän yksittäisten komponenttien vaikutusta haun tuloksiin yksinkertaistetussa asetelmassa. Mallin rajoituksia ovat hakuaiheiden, dokumenttikokoelman sekä relevanssiarvioiden rajallisuus ja malliin sisältyvät epärealistisen yksinkertaistavat oletukset, esimerkiksi relevanssin käsitteen redusoinen informaation ”aiheenmukaisuudeksi”.¹⁸

Relevanssi

Tiedonhaun tausta-ajatuksena on kulloiseenkin inhimilliseen tiedontarpeeseen sopivan tiedon tarjoaminen. Tiedonhaun tutkimus pyrkii yhä parempien tiedonhakujärjestelmien kehittämiseen. Sen määrittely, mikä on parempi järjestelmä, riippuu tarkastelukulmasta ja tutkimusasetelmasta. Tämä paremmuuden tarkastelu tarkoittaa

17 Ks. esim. Arvola ym. 2010; Baskaya ym. 2011; Keskustalo ym. 2009; Pääkkönen ym. 2017.

18 Lisätietoa mallin rajoitteista tarjoavat Turpin & Scholer 2006 ja Smith & Kantor 2008.



Kuvio 4. Relevanssin tyypit ja erilaisia tutkimusnäkökulmia.

tiedonhaun tuloksellisuuden eli onnistumisen tai epäonnistumisen tarkastelua. Tästä tarkastelusta puhutaan *tiedonhaun evaluointina*, jolloin jollain keinolla pyritään osoittamaan se, kuinka “hyvin” tiedonhaku toimii suhteessa tarkastelukulmaan ja tarkastelun kohteeseen. Tarkastelunäkökulmia voidaan rajata esimerkiksi aiemmin esitettyjen kolmen käyttäjäasetelman mukaisesti sekä tarkastelukohteen eli tiedonhaun osatekijöiden mukaan (kuvio 4).

Olennessa käsite tiedonhaun evaluoinnissa on *relevanssi*. Relevanssi tarkoittaa arviota informaation jonkin tyyppisestä “käyttökelpoisuudesta” jossain tilanteessa eli sitä, miten hyvin haettu tieto vastaa tiedontarpeeseen. Toisin sanoen relevanssi on suhde tiedontarpeen ja tiedon välillä.¹⁹ Relevanssin määrittelyn tekee vaikeaksi se, että tiedontarpeen käsitteen määrittely vaihtelee eri tutkimuk-

19 Rieh & Xie 2006; Saracevic 1975.

sisä.²⁰ Relevanssi voidaan jakaa karkeasti *järjestelmien relevanssiin* ja *käyttäjärelevanssiin*.²¹ Nämä esittävät ilmiöitä ikään kuin kahdessa eri maailmassa – *järjestelmien maailmassa* ja *käyttäjän (tiedon käyttäjän) maailmassa* – jotka ovat vuorovaikutuksessa keskenään. Puhutaankin *järjestelmälähtöisestä* ja *käyttäjälähtöisestä* tiedonhaun tutkimuksesta.

Kuviossa 4 on esitetty monipuolisempi jaottelu, jossa lähdetään algoritmista relevanssista kohti tiedon hakijaa ja hakijan toiminnan konteksteja. Voidaan ajatella, että siirryttäessä kuviossa vasemmalta oikealle tarkastelualue laajenee huomioimaan kasvavassa määrin inhimillisen kontekstin ulottuvuuksia. Relevanssista löytyy myös muita esityksiä ja jäsennyksiä.²²

Kuvioon on merkitty myös tutkimustraditioita, joilla on erilainen näkökulma ilmiöihin. Jatkumon oikeassa ääripäässä tarkastellaan todellisen maailman ilmiöitä sellaisina kuin ne esiintyvät yhteisöissä, vasemmalla puolestaan on kuvattu kokeellinen ja voimakkaasti rajattu tiedonhaun laboratorioasetelma. Tällöin inhimillinen ulottuvuus ja ihmisen taustan huomiointi vähenee ja relevanssin määritelmä vaihtuu ja liukuu toisenlaiseksi.²³

Järjestelmä- eli *algoritminen relevanssi* tarkoittaa tiedonhakujärjestelmälle annetun *kyselyn* ja *dokumenttien esitysten* – käytännössä indeksien eli hakemistojen – välistä suhdetta eli sitä, miten merkijonoja täsmäytetään järjestelmässä ja onnistutaanko siinä (täsmäytyksestä, ks. luku 3.1) Jokaisella järjestelmällä on keinonsa esittää dokumentit, järjestää ne ja täsmäyttää eli verrata niitä kyselyyn. Tässä siis tarkastellaan sitä, miten järjestelmä päättelee sen, liittyykö dokumentin informaatio annettuun kyselyyn.

Aiherelevanssi on varsin suosittu käsitteellistys tiedonhakujärjestelmiä tarkasteltaessa, jolloin tiedonhakijan tilanteen ja taustan

20 Byström & Kumpulainen 2020.

21 Saracevic 2016.

22 Ks. esim. Belkin 2015; Borlund 2003; Cosijn & Ingwersen 2000; Ingwersen & Järvelin 2005; Saracevic 2007.

23 Kelly 2009.

tarkastelun osuus on kuvattu tutkimusasetelmassa vain hakutehtävän muodossa – joko simuloidun tilanteen kuvauksena²⁴ tai pelkistettynä hakuaiheeksi. Tämä tuo hallittavuutta ja toistettavuutta mittaamiseen, kun todellisen maailman monimuotoiset hakuaiheet ja hakutarpeet pelkistetään kontrolloiduiksi tehtäviksi, jotka toistuvat eri tutkimuksissa ja mittauksissa samoina. Aiherelevanssin mielessä haettu tieto (dokumentti) on relevanttia, jos se vastaa hakutehtävässä määriteltyä *aihetta*. Aiherelevanssin arviointi tapahtuu vertaamalla dokumentin informaatioisältöä hakuaiheeseen tai hakutehtävään. Tällainen dokumentin tai informaation relevanssi on suhteellisen pysyvä ominaisuus. Esimerkiksi jos hakuaiheeksi on kuvattu ”Etsi tieto siitä, milloin Tampereen kaupunki on perustettu”, relevantti dokumentti kertoo vastauksen tähän – käyttötilannetta ei tarvitse ottaa huomioon relevanssin tarkastelussa. Aiherelevanssi on varsin suosittu käsitteellistys *hakujärjestelmiä* tarkasteltaessa. Sen etuna on, että tiedonhakijan tilanteen ja taustan tarkastelun osuus kutistuvat tutkimusasetelmassa. Aiherelevanssi määritellään arvioimalla, *sisältääkö dokumentti informaatiota hakuaiheessa määritellystä aihepiiristä*. Aiherelevanssi arvioi vain dokumentin aiheenmuokaisuutta, eikä sen arvioinnissa huomioida käyttäjien subjektiivisia tarpeita.

Käyttäjärelevanssi tarkoittaa tiedon hakijan *tiedontarpeisiin* vastaavaa informaatiota. Tässä voidaan ottaa huomioon joukko erilaisia piirteitä, jotka liittyvät esimerkiksi tiedonhakijan tietämyksen tilaan (*kognitiivinen relevanssi*), tunteisiin ja motivaatioon (*motivaatiorelevanssi*) tai hakutilanteeseen tai -tehtävään (*tilanne- ja tehtävärelevanssi*). *Hakijan tietämyksen tilan* piirteitä ovat aiempi, jo hankittu tieto sekä dokumentin tulkinta- ja tiedonhakutaidot. Tunteet ja motivaatio vaikuttavat myös siihen, kuinka informaatiota arvioidaan suhteessa tavoitteisiin ja millaista tietoa hyväksytään tai preferoidaan. Relevanssi on tiedon hakijan näkökulmasta katsottuna dynaamista, sillä tilanteesta ja tehtävästä opitaan tiedonhaun aikana.

24 Borlund 2003; 2016.

Samalla ymmärrys tiedontarpeesta muuttuu tiedon karttumisen myötä. Tiedontarpeiden katsotaan tilanne- ja tehtävärelevanssissa kumpuavan toiminnan kehyksestä eli tilanteen tai tehtävän kulloisestakin tulkinnasta. Lisäksi tilannesidonnaisuus määrittää tiedonhakua, jolloin esimerkiksi tilanteen ollessa ohi tiedontarve häviää ja haetun dokumentin relevanssi muuttuu. Tällainen tilanne voi olla esimerkiksi linja-auton aikataulun hakeminen – kun tarve on ohi, ei informaatiolla ole enää merkitystä tiedonhakijalle.

Sosiaalisesta relevanssista puhuttaessa tarkastellaan sitä, kuinka informaatio sekä informaatio- ja tietojärjestelmät vaikuttavat työhön, vapaa-aikaan, yhteiskuntaan ja kulttuuriin.²⁵ Sosiaalisen median alustat voidaan nähdä hakukoneina, ja niillä on ollut suuri vaikutus julkiseen keskusteluun ja yhteisöjen viestinnän muutokseen. Esimerkiksi COVID-19-epidemian aikana vuosina 2020–2021 liikkui runsaasti dis- ja misinformaatiota, joka ei ollut valtaväestön ja viranomaisten mielestä sopivaa tai informaation katsottiin olevan jopa haitallista yleiselle terveydelle. Tällä tiedolla oli vaikutusta erilaisiin yhteisöihin ja niiden yhteisöjen toimintaan, mutta se myös aiheutti paljon julkista keskustelua. Normatiivinen tieto tarkoittaa sitä, että tieto vastaa yhteiskunnassa vallitsevia normeja ja arvoja. Sosiaalisen relevanssin yhteydessä voidaan katsoa, että tieto on relevanttia, jos se on yleisesti hyväksyttyä ja hyväksytyjen auktoriteettien hyväksymää.

Tässä luvussa kuvasimme tiedonhakua ilmiönä, jonka ytimessä on pyrkimys yhdistää hakijan tiedontarpeen kuvauksia ja dokumenttien tietosisällön kuvauksia hyödyllisen tai relevantin informaation löytämiseksi. Korostimme artikkelissa näkemystä, että *todelliset* hakijat ja hakemisen käyttöyhteydet ovat moninaisia ja *tutkimuksen* aikana hakuprosessia ja hakijaa aina jotenkin *mallinnetaan*. Mallinnus sisältää valintoja ja on siksi lähtökohtaisesti rajaavaa. Mallinnus ei ole ongelma, josta tulisi päästä eroon, mutta valitun mallinnuksen rajoista olisi hyvä tulla tietoiseksi. Valaisimme asiaa

25 Saracevic 1995.

esittämällä jaon kolmentyyppisiin tiedonhaun tutkimuksiin. Ensimmäisessä tutkimustyyppissä tutkitaan ihmisten tiedonhakutoimintaa *autenttisissa* tilanteissa. Toinen tutkimuksen teon tapa hyödyntää koehenkilöitä, jotka suorittavat heille *annettuja* tehtäviä, jotka saattavat muistuttaa heille tuttuja tehtäviä tai olla enemmän tai vähemmän vieraita. Kolmanneksi tiedonhakua voidaan tutkia vielä edellisiä pelkistetyemmin testikokoelmaa käyttäen, jopa kokonaan *ilman* ihmisten läsnäoloa. Testikokoelman kolme peruskomponenttia ovat *hakuaiheet*, *dokumenttikokoelma* sekä *relevanssiarviot*. Kukin näistä kolmesta tutkimuksenteon tavasta ottaa omilla tavoillaan kantaa siihen, millaisena tutkimusasetelmaan sisältyvä tiedonhakija tutkimuksessa näyttäytyy eli minkä piirteiden kautta hakija tulee tutkimuksessa mallinnetuksi.

Lähteet

- Arvola, Paavo & Tuulikki Alamettälä (2022a) IRVILAB: Gamified searching on multilingual Wikipedia. *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, 3329–3333.
- Arvola, Paavo & Tuulikki Alamettälä (2022b) Enhancing information literacy skills: A game design for seeking information and making queries. *Proceedings of games and learning alliance. GALA 2022. Lecture notes in computer science* 13647, Cham: Springer, 323–328.
- Arvola, Paavo & Jaana Kekäläinen & Marko Junkkari (2010) Expected reading effort in focused IR evaluation. *Information retrieval* 15:5, 460–484.
- Baskaya, Feza & Heikki Keskustalo & Kalervo Järvelin (2011) Simulating simple and fallible relevance feedback. *Proceedings of the 33rd European conference on advances in information retrieval*, 593–604.
- Belkin, Nicholas J. (2015) Salton award lecture: people, interacting with information. *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, 1–2. <https://doi.org/10.1145/2766462.2767854>
- Borlund, Pia (2003) The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information research* 8:3, 3–8.
- Borlund, Pia (2016) A study of the use of simulated work task situations in interactive information retrieval evaluations: A meta-evaluation. *Journal of documentation* 72:3, 394–413.
- Byström, Katriina & Kalervo Järvelin (1995) Task complexity affects information seeking and use. *Information processing & management* 31:2, 191–213.

- Byström, Katriina & Jannica Heinström & Ian Ruthven (toim.) (2019) *Information at work: Information management in the workplace*. Lontoo: Facet.
- Byström, Katriina & Sanna Kumpulainen (2020) Vertical and horizontal relationships amongst task-based information needs. *Information processing & management* 57:2, 102065. <https://doi.org/10.1016/J.JIPM.2019.102065>
- Cleverdon, Cyril W. (1991) The significance of the Cranfield tests on index languages. *Proceedings of the 14th annual international ACM SIGIR conference on research and development in information retrieval*, 3–12.
- Cosijn, Erica & Peter Ingwersen (2000) Dimensions of relevance. *Information processing & management* 36:4, 533–550. [https://doi.org/10.1016/S0306-4573\(99\)00072-2](https://doi.org/10.1016/S0306-4573(99)00072-2)
- Huuskonen, Salla & Pertti Vakkari (2010) Client information system as an everyday information tool in child protection work. *Proceedings of the third symposium on information interaction in context*, 3–12.
- Ingwersen, Peter & Järvelin Kalervo (2005) *The Turn: Integration of information seeking and retrieval in context (The information retrieval series)*. Dordrecht: Springer.
- Korkeamäki, Laura & Sanna Kumpulainen (2019) Interacting with digital documents: A real life study of historians' task processes, actions and goals. *CHIIR 2019 - Proceedings of the 2019 conference on human information interaction and retrieval*. <https://doi.org/10.1145/3295750.3298931>
- Kelly, Diane (2009) Methods for evaluating interactive information retrieval systems with users. *Foundations and trends in information retrieval* 3:1—2, 1–224.
- Keskustalo, Heikki & Kalervo Järvelin & Ari Pirkola & Tarun Sharma & Marianne Lykke (2009) Test collection-based IR evaluation needs extension towards sessions – A case of extremely short queries. *Information retrieval technology: The 5th Asia information retrieval symposium (AIRS'09)*. LNSC 5839, 63–74.
- Kettunen, Kimmo & Heikki Keskustalo & Sanna Kumpulainen & Tuula Pääkkönen & Juha Rautiainen (2022) OCR quality affects perceived usefulness of historical newspaper clippings. A user study. Teoksessa Giorgio Maria di Nunzio & Beatrice Portelli & Dominico Redavid & Gianmaria Silvello (toim.) *Proceedings of the 18th Italian research conference on digital libraries (CEUR workshop proceedings 3160)*. CEUR-WS.org. <http://ceur-ws.org/Vol-3160/paper1.pdf>
- Kumpulainen, Sanna (2013) *Task-based information access in molecular medicine: Task performance, barriers, and searching within a heterogeneous information environment*. Väitöskirja. Tampere: Tampereen yliopisto. <http://urn.fi/URN:ISBN:978-951-44-9287-7>
- Kumpulainen, Sanna (2017) Task-based information searching. Teoksessa John D. McDonald & Michael Levine-Clark (toim.) *Encyclopedia of library and information sciences*. Boca Raton: CRC Press, 4526-4536. <https://doi.org/https://doi.org/10.1081/E-ELIS4>
- Kumpulainen, Sanna & Kalervo Järvelin (2010) Information interaction in molecular medicine: Integrated use of multiple channels. *IliX 2010* -

- Proceedings of the 2010 information interaction in context symposium*. <https://doi.org/10.1145/1840784.1840800>
- Kumpulainen, Sanna & Heikki Keskustalo & Boyang Zhang & Kostas Stefanidis (2020) Historical reasoning in authentic research tasks: Mapping cognitive and document spaces. *Journal of the association for information science and technology* 71:2, 230–241. <https://doi.org/10.1002/asi.24216>
- Odijk, Daan & Ryen White & Ahmed Hassan Awadallah & Susan Dumais (2015) Struggling and success in web search. *Proceedings of the 24th ACM international on conference on information and knowledge management*, 1551–1560.
- Pääkkönen, Teemu & Jaana Kekäläinen & Heikki Keskustalo & Leif Azzopardi & David Maxwell & Kalervo Järvelin (2017) Validating simulated interaction for retrieval evaluation. *Information retrieval journal* 20:4, 338–362.
- Rieh, Soo Young & Hong Iris Xie (2006) Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information processing & management* 42:3, 751–768.
- Saastamoinen, Miamaria & Kalervo Järvelin (2017) Search task features in work tasks of varying types and complexity. *Journal of the association for information science and technology* 68:5, 1111–1123. <https://doi.org/10.1002/asi.23766>
- Saastamoinen, Miamaria & Sanna Kumpulainen (2014) Expected and materialised information source use by municipal officials: Intertwining with task complexity. *Information research* 19:4. <http://InformationR.net/ir/19-4/paper646.html>
- Salton, Gerard & Michael J. McGill (1987) *Introduction to modern information retrieval*. Auckland: McGraw-Hill.
- Saracevic, Tefko (1975) Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American society for information science* 26:6, 321–343.
- Saracevic, Tefko (1995) Evaluation of evaluation in information retrieval. *Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval*, 138–146. <https://doi.org/10.1145/215206.215351>
- Saracevic, Tefko (2007) Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American society for information science and technology* 51:13, 1915–1933. <https://doi.org/10.1002/asi.20682>
- Saracevic, Tefko (2016) The notion of relevance in information science: Everybody knows what relevance is. But, what is it really? *Synthesis lectures on information concepts, retrieval, and services* 8, i–109.
- Sormunen, Eero & Sakari Hokkanen & Petteri Kangaslampi & Petri Pyy & Bemmu Sepponen (2002) Query performance analyser - A web-based tool for IR research and instruction. *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval*, 450–450.
- Sormunen, Eero & Juha Laaksonen & Heikki Keskustalo & Jaana Kekäläinen & Harri Kempainen & Harri Laitinen & Ari Pirkola & Kalervo Järvelin (1998) IR-game - a tool for rapid query analysis in cross-language IR

- experiments. *Proceedings of the joint workshop on cross language issues in artificial intelligence & issues of cross cultural communication*, 24th November, 1998, Kent Ridge Digital Labs, Singapore. <https://jsaic.krdl.org.sg/pricaia98>
- Smith, Catherine L. & Paul Kantor (2008) User adaptation: Good results from poor systems. *Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval*, 147–154.
- Turpin, Andrew & Falk Scholer (2006) User performance versus precision measures for simple search tasks. *Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval*, 11–18.
- Webber, Bonnie (2010) Question answering. Teoksessa Alexander Clark & Chris Fox & Shalom Lappin (toim.) *The handbook of computational linguistics and natural language processing*, 630–654. West Sussex: Blackwell Publishing Ltd.
- White, Ryan W. & Jeff Huang (2010) Assessing the scenic route: measuring the value of search trails in web logs. *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, 587–594. <https://doi.org/10.1145/1835449.1835548>
- Wolfram, Stephen (2023) *What is ChatGPT doing ... and why does it work?* Champaign: Wolfram Media.
- Voorhees, Ellen M. (2007) TREC: Continuing information retrieval's tradition of experimentation. *Communications of the ACM*. 50:11, 51–54.
- Voorhees, Ellen M. & Dawn M. Tice (2000) Building a question answering test collection. *Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval*, 200–207. New York: Association for Computing Machinery. <https://doi.org/10.1145/345508.345577>
- Yu, Fangyi & Lee Quartey & Frank Schilder (2023) Exploring the effectiveness of prompt engineering for legal reasoning tasks. *Findings of the association for computational linguistics*: ACL 2023, 13582–13596. Toronto: Association for Computation.

3.3 Algoritmit ja tiedon löydettävyys

Sanna Talja

- Hakukoneiden, sosiaalisen median, musiikin ja elokuvien striimausalustojen sekä monien muiden verkkopalvelujen ja teknologisten alustojen toiminta perustuu algoritmeihin.
- Algoritmit ovat reseptejä, joiden kautta meille suodattuu tuloksia Google-hauissa tai musiikkisuosituksia Spotifyssa.
- Algoritmit puitteistavat tiedonsaantia ja luovat mahdollisuuksia tietämiselle ja viestinnälle.
- Algoritmit eivät ole vain neutraaleja työkaluja tai välineitä, sillä näkymättömänä taustana ja verkon infrastruktuurina ne myös monella tapaa muovaavat ja rajaavat mahdollisuuksia hakutehtävien toteuttamiseen ja halutun tiedon löytämiseen.

Tässä artikkelissa tarkastellaan sitä, kuinka tiedonsaannin mahdollisuudet digitaalisessa ympäristössä muotoutuvat ja puitteistuvat. Tarkemmin käsitellään seuraavia kysymyksiä: Mitä algoritmit ovat? Miten ne toimivat? Mitä tiedämme algoritmista suodattamisesta? Lähemmin tarkastellaan haku- ja suositusalgoritmeja sekä niihin läheisesti liittyvää käyttäjien autonomian ja seurannan ongelmakenttää. Sosiotekninen algoritmitutkimus on monitieteinen tutkimusalue, johon tiedon löydettävyyden ongelmakenttään keskittyvä informaatiotutkimus osaltaan kontribuoi. Informaatiotutkimuksen perinteessä keskeisellä sijalla olleesta hakukoneiden ja tiedonhaun tutkimuksesta kriittinen algoritmitutkimus eroaa siinä, että se puureutuu algoritmien yhteiskunnallisiin, poliittisiin ja eettisiin ulottuvuuksiin. Sosioteknisestä algoritmitutkimuksesta voi luonnollisesti olla paljonkin tukea algoritmien kehittämisessä, mutta varsinainen

tiedonhaku- tai suositusalgoritmien luominen ei kuulu kriittisen algoritmitutkimuksen piiriin.

Mitä algoritmit ovat?

Algoritmit ovat menettelystrategioita ja ohjesääntöjä. Ratkaistava ongelma puretaan pieniin osiin tai askeleisiin, joiden kautta annettujen sääntöjen mukaan edeten saavutetaan toivottu lopputulos.¹ Askel askeleelta säännön mukaan etenevä ongelmanratkaisu sopii luonnollisesti vain tarkoin määritelyihin ja eksaktisti määriteltäviin ongelmiin.² Algoritmeista voidaan puhua myös tietokoneohjelmina, vaikka algoritmit voivat olla olemassa myös pelkästään matemaattisina kaavoina. Algoritmin toiminta perustuu syötteiden – useimmiten datan – laskennalliseen käsittelyyn.³

Digitaalisessa ympäristössä tärkeitä algoritmeja käyttäjän näkökulmasta ovat:

- Hakualgoritmit eli hakukonealgoritmit (esimerkiksi Googlessa on oma algoritminsä, Facebookissa ja X:ssä omansa)
- Suositusalgoritmit (esimerkiksi musiikin suositellu Spotifyssa, ohjelmasuosittukset Netflixissä tai videosuosittukset YouTubeissa)
- Suodatusalgoritmit (esimerkiksi uutissyötteiden suodatus Facebookissa tai Instagramissa)
- Verkkokäyttäytymisen perusteella personoidut mainokset

Hakualgoritmit

Ensimmäiset hakukoneet kuten Altavista näyttivät hakutuloksina yksinkertaisesti kaikki sivut, jotka vastasivat käyttäjän hakukonee-

1 Dutta & Patel 2022.

2 Dutta & Patel 2022.

3 Gillespie 2014.

seen syöttämiä avainsanoja. Hakulistaussijoitukseen vaikutti se, kuinka monta kertaa haetut sanat esiintyivät sivulla, esiintyivätkö sanat sivun otsikossa sekä miten lähellä toisiaan haetut sanat olivat sivulla. Varhaiset hakukoneet tuottivat runsaasti epärelevantteja tuloksia. Edistyneemmät haut hakusanoja yhdistelemällä olivat mahdollisia, mutta käyttäjille hankalia.

Googlen alkuperäinen hakualgoritmi *PageRank* julkaistiin vuonna 1996. Sen jälkeen hakualgoritmiin on kuitenkin tehty kymmeniätuhansia suurempia ja pienempiä muutoksia. Tällä hetkellä Googlen osuus tehdyistä hauista on noin 86–90 prosenttia maasta ja laitteesta riippuen. Muiden hakukoneiden kuten Bingin ja Yagoon osuus on vain prosentteja (molemmat alle 5 %). Monet hakukoneet, kuten Startpage, hyödyntävät Googlen hakemistoa, joten ne ovat vain näennäisiä vaihtoehtoja Googlelle.⁴ Hakemisto syntyy, kun erityinen ohjelma eli hakurobotti seuraa linkkejä ja lisää tiedon hakemistoon. Ilman hakemistoa ei olisi mahdollista käydä sivuja ja niiden sisältöä läpi riittävän nopeasti. Se, mitä tietoa hakemistoon lisätään, riippuu hakukoneen indeksointikriteereistä.⁵ Eri hakukoneita käyttämällä ei saada juurikaan erilaisia hakutuloksia. Hakukoneiden välillä on eroja siinä, missä määrin käyttäjien yksityisyyttä suojellaan. Kaikki hakukoneet eivät kerää käyttäjien hakuhistoriaa tai myy sitä mainostajille.

Googlen ylivoimaisuus perustuu siihen, että hakutulokset perustuvat käyttäjän hakukoneeseen syöttämiin hakusanoihin ja -fraaseihin, mutta Googlen hakualgoritmi mittaa myös sivujen tärkeyttä ja keskeisyyttä. PageRank toimii siten, että se laskee sivulle vieviä linkkejä, joten jokainen sivulle johtava linkki on ikään kuin kannatusääni.⁶ Mitä enemmän linkkejä, sitä korkeammalle sivusto kohoaa PageRankissa. Pelkällä linkkien määrällä ei ole merkitystä, vaan myös niiden laadulla. Tärkeiltä sivuilta tulevilla linkeillä on enemmän pai-

4 Ks. Lekander 2025.

5 <https://www.google.com/search/howsearchworks/how-search-works/organizing-information/>

6 <https://fi.wikipedia.org/wiki/PageRank>

noarvoa kuin vähemmän tärkeiden sivujen linkeillä. Jokainen sivu tulee rankatuksi Googlen omien arviointikriteerien perusteella.⁷ Ensisijaisia kriteerejä ovat arvovaltaisuus ja hyödyllisyys. Googlen tekoäly ei siis analysoi pelkästään yksittäisiä sanoja, vaan pyrkii ymmärtämään kunkin sivuston luonnetta ja kontekstia.

Googlen algoritmissa on 210 eri luokittelutekijää, joiden mukaan sivut järjestyvät hakutuloksissa.⁸ Sivuston julkaisijan ja käyttäjän kannalta on tietenkin ensiarvoisen tärkeää, mitkä sivut asetuvat hakutuloslistauksessa ylimmäksi. Harva käyttäjä etenee tarkastelemaan hakutuloslistauksessa alemmas, esimerkiksi toiselle sivulle jääviä linkkejä.⁹ Kriteereinä näkyvyydelle ja sijoittumiselle hakutuloslistauksessa ovat tiedon tuottajan arvovaltaisuus eli auktoriteettiasema ja sisällön alkuperäisyys. Googlen hakutuloksissa arvovaltaisia eli tuloslistalle ensimmäiseksi sijoitettuja lähteitä ovat Wikipedia, isot uutistoimistot sekä isot perinteiset sanomalehdet ja uutiskanavat, kuten YLE Suomessa.¹⁰

Monet hakukoneita pitkään käyttäneet katsovat, että Google-haut tuottivat alkuvaiheessa laajempia ja monipuolisempia tuloksia. Kone esitti tuloksia vapaammin eri puolelta verkkoa hakusanojen perusteella. Rankkauskriteerien kiristäminen ja pyrkimys niin sanotun valetiedon tai kyseenalaisten sivustojen torjumiseen merkitsee sitä, että Google tekee relevanssiarviot käyttäjän puolesta. Sivustojen ja sisältöjen tarkastamisessa käytetään tekoälyn lisäksi inhimillisiä sisältö tarkastajia.¹¹ Mikäli tekstissä tai puheessa esiintyy tarkastettava ja kontrolloitava sana tai ilmaus, inhimilliset tarkastajat ottavat sivustontuottajan lähempään tarkasteluun ja varoittavat mahdollisista käytösaäntöjen rikkomuksista. Viime kädessä sisältötuottajan profiili poistetaan kokonaan alustalta.¹² Yleensä profiili

7 Google 2022.

8 Rasmussen 2021.

9 Dean 2022.

10 Lundahl 2022.

11 Langvardt 2017.

12 Shir-Raz ym. 2022.

lien poisto tapahtuu yhtäaikaaisesti yli alustojen (Facebook, Instagram ja YouTube) ja mahdollisesti myös rahoitusalustoilta (Patreon ja Paypal), mikäli kyseessä on yleisörahoitteinen vaihtoehtojournalismi.

Alkuvaiheessa (2009–2016) Googlen ja PageRank-algoritmin tavoitteena oli siis tarkkuus ja kattavuus suhteessa hakusanoihin. Rankkausperiaate oli puhtaammin matemaattinen eli Google ensisijaisti sivut, joihin oli eniten linkkejä. Hakija arvioi itse lähteiden osuvuuden ja etsi itse lähteistä vastauksen. Hakutulokset olivat lähdeprefenssien osalta neutraaleja eli käyttäjiä ohjaamattomia. Google oli luonteeltaan sosiaalinen verkosto, joka yhdisti erilaisia näkökulmia ja tuloksia yhteen paikkaan hakutulosten osalta.

Googlen relevanssiarvioinnin kehitykseen vaikuttavat sekä poliittiset että tekniset syyt. Sivustolle johtavien linkkien määrä osoittautui liian helposti manipuloitavaksi, joten matemaattinen äänestysperiaate ei toiminut enää relevanssiarvioiden pohjana. Googlen filosofiassa tapahtui myös käänne, sillä nykyisin pyrkimyksenä on antaa käyttäjälle hakuun yksi oikea ja tarkka vastaus mahdollisimman pienellä vaivalla. Monesti käyttäjät saavatkin vastauksen hakuunsa suoraan hakutulossivulta ilman, että heidän tarvitsee edetä klikkaamaan tuloslistalla olevia linkkejä.¹³ Hakutulossivulla on nos-tona vastausikkuna (*answer box*, *information box*) joka usein on Wikipedia-sivu tai terveyteen liittyvissä hauissa lääketieteen perusteos.

Terveystiedon osalta on tärkeää huomioida, että Googlen emoyhtiö Alphabet Inc sisältää myös lääketieteellisuuden haaroja. Vaihtoehto- tai luonnonlääketiedettä edustavat aiheet tai sivut joutuvat helposti varjobannatuiksi (löydettävyyden ja näkyvyyden rajoittaminen) tai sivustojen mustalle listalle.¹⁴

Googlen sivustorankkaus vaikuttaa siten, että käyttäjän oma-kohtainen mahdollisuus sivustojen ja niiden sisältämän tiedon ar-

13 Tung 2019.

14 Brownstein 2019; Goldman 2019; MacLean 2017; Shir-Raz ym. 2022. Googlen mustista listoista tarkemmin Grind ym. 2019; Epstein 2016.

viointiin on vähentynyt.¹⁵ Koska vähemmän arvovaltaiset sivustot, esimerkiksi kansalaisliikkeet, yhteisöt ja yksittäiset henkilöt, kuten tutkivat journalistit, jotka eivät ole osa isoja mediataloja, jäävät hakutuloslistauksessa alas, olemassa olevan tiedon ja keskustelujen yksityiskohtaisempi tarkastelu on huomattavasti hankalampaa kuin ennen vuotta 2016. Tässä mielessä Googlen rooli on muuttunut hakukoneesta ja -alustasta enemmänkin julkaisijaksi.¹⁶ Perinnemedia ja Googlen hakualgoritmi toimivat samalla tavoin uutisvirran suodattajina. Perinnemedian melko yhtenäinen tapa valita ja esittää ajankohtaisia sisältöjä ja näkökulmia aiheisiin heijastuu myös hakutuloksiin.¹⁷

Googlen tiukat laatuksiteerit ovat käyttäjän etu siinä mielessä, että nykyisin ei esimerkiksi voi vahingossa päätyä aikuisviihdesivustolle etsiessään jotakin muuta. Toisaalta asioihin perehtymisen mahdollisuus on kaventunut, koska niin suuri osa vähemmän tunnetuista lähdevaihtoehdoista asettuu tuloslistalla kauas kärjestä tai karsiutuu automaattisesti ulos hakutuloksista. Emme tiedä, millaista tietoa emme löydä hakutuloslistalta, koska saamme joka tapauksessa tuloksia. Tässä mielessä hakukonealgoritmit asettavat tietämissen rajoja.¹⁸

Suosittelualgoritmit

Suositusjärjestelmät ovat sekä suosittelu- että suodatusjärjestelmiä. Kautta aikojen ihmiset ovat halunneet kuulla muiden luotettujen vinkkejä siitä, mikä on esimerkiksi hyvää ja lukemisen arvoista kirjallisuutta tai mikä olisi viihdyttävä ja koukuttava tv-sarja. Ennen verkkokirjastojen kehittymistä ihmiset lainasivat enemmän kirjoja kirjastojen palautushyllyistä, koska se, että muut olivat lainanneet kirjan, oli jonkinasteinen merkki suosituksesta. Automaattiset tois-

¹⁵ Grind ym. 2019.

¹⁶ Lundahl 2022.

¹⁷ Lundahl 2022, 1449.

¹⁸ Gillespie 2014.

ten käyttäjien valintoihin perustuvat suositukset tulivat ensimmäisenä käyttöön Amazon-verkkokirjakaupassa. Amazonin suositusalgoritmin varhaisin versio suositteli käyttäjälle hänen valintojensa perusteella aineistoa, jota muut samasta aineistosta kiinnostuneet käyttäjät olivat ostaneet. Nämä suositukset palvelivat uusien suosikkien löytämistä käyttäjän tietoisten ja piilevien mieltymysten pohjalta.

Yhteistoiminnalliset eli kollaboratiiviset suositukset perustuvat siihen, että järjestelmä vertaa käyttäjän antamaa tai hänestä kerättyä tietoa järjestelmässä olevaan dataan, laskee yhtäläisyyksiä käyttäjien välillä ja ennustaa vertailun perusteella, miten sopivana käyttäjä mitäkin kohdetta pitäisi. Käyttäjä löytää itselleen sopivia kohteita, joita ei muulla tavoin voisi löytää. Samankaltaisten käyttäjien mieltymyksiin nojaavat suositukset ovat parhaimmillaan hyvinkin osuvia, ja ne ovat olleet valtava edistysaskel tiedonhaun helpottamisessa. Esimerkiksi Spotifyn kykyä suositella sisältöjä, joista käyttäjä saattaisi aiempien valintojensa perusteella pitää ja josta hän ei entuudestaan ole ollut tietoinen, on kehuttu jopa ”maagisen hyväksi”.¹⁹

Yhteistoiminnallisissa suositusjärjestelmissä keskeinen toimintaperiaate on ihmisryhmien, joilla on samankaltaisia intressejä, ja samankaltaisten sisältöjen identifiointi. Pääperiaatteena on tarjota käyttäjälle sisältöjä hänen aikaisempiin valintoihinsa ja preferensseihinsä perustuen. Yhteistoiminnallisten suositusten erityinen vahvuus on se, että ne palvelevat myös niitä, joilla on erityisiä intressejä ja jotka harrastavat vähemmän tunnettua musiikkia.

Useimmat alustat kuten Spotify, Netflix ja YouTube käyttävät kuitenkin useiden erilaisten suositusalgoritmien ja -perusteiden yhdistelmää. *Sisältöperustaiset* suositukset perustuvat sisältöjen tarkkaan ja hienovaraiseen etukäteiskuvailuun, ja niiden toimintaperiaate muistuttaa perinteistä tiedon organisointia, kuten kirjastojen luokitusjärjestelmiä. Luokitteluperusteita elokuvien ja tv-sarjojen osalta ovat esimerkiksi genre, tunnelma ja maanosa. Asiantuntijat

19 Hu 2021; Jones 2022.

tekevät kuvailut tarkasti ja kattavasti etukäteen. Esimerkiksi Netflixissä on noin 20 pääluokkaa ja 4 000 erilaista alagenreä.²⁰ Tarkan genreluokituksen olemassaolosta huolimatta Netflix ja Spotify eivät tee sisältöjään suoraan käyttäjän haettavaksi genren tai genrekoodien perusteella.²¹ Vaikka tämä mahdollisuus löytyy palvelusta, suositusten ensisijaisena perusteena ovat käyttäjäprofiilin lisäksi sisältöjen suosituimmuus ja ajankohtaisuus. Yhteistoiminnallisen ja sisältöperustaisen suosittelemisen lisäksi käytetään *demografista profiilointia* eli suosituksia tarjotaan iän, sukupuolen ja maantieteellisen sijainnin perusteella. Useimpien alustojen algoritmit osaavat päätellä nämä seikat ilman käyttäjän itse antamaa informaatiota.

Yleensä omien aikaisempien hakujen ja käyntikohteiden oletetaan olevan suositusten pääperiaate. Oletuksena on, että sisällöt, joista käyttäjä ei ole aiemmin ollut kiinnostunut, jäävät piiloon tai suositusten ulkopuolelle. Algoritmien toiminta ja sisältösuodatus ei kuitenkaan ole täysin personoitua. Algoritmit vaihtelevat alustakohtaisesti, ja painotukset suodatusperiaatteissa ovat kehittyneet ajan myötä. Emme voi tietää tarkalleen, mitkä suodatus- ja suositusperiaatteet milloinkin ja milläkin alustalla ovat johtavia ja mitkä toissijaisia. Luonnollisesti kuitenkin huomaamme, mikäli meille tarjotaan vaikkapa YouTubessa runsaasti ajankohtaisia sisältöjä, joita emme koskaan ole hakeneet ja jotka eivät lainkaan vastaa normaaleja kiinnostuksen kohteitamme. Huomaamme, mikäli äsken hakukoneen kautta etsimämme aihe tai tuote ilmenee uutissyötteessämme esimerkiksi Facebookissa kohdennettuina mainoksina. Näinä hetkinä algoritmit tulevat näkyviksi käyttäjille, kun ne yleensä toimivat enemmänkin näkymättöminä, taustalla.²²

20 Ks. <https://www.netflix-codes.com>

21 Dershowitz 2024.

22 Dogruel ym. 2022.

Uutissyötteet, suositukset ja käyttäjän kontrolli

Käyttäjät eivät pysty itse täysin kontrolloimaan suositusperusteita, esimerkiksi sulkemaan pois demografisia suosituksia, trendaavia uutisia tai ei-kiinnostavia aihealueita. Käyttäjän ja algoritmien valtasuhde on siis epätasapainoinen.²³ Käyttäjän mahdollisuudet vaikuttaa algoritmien luomiin informaatiovirtoihin ja ympäristöihin ovat rajalliset.

Erityisen suuri ongelma käyttäjän kontrollin näkökulmasta on, että alustat saattavat painottaa suosituksia ja suositustoimintoa jopa hakuja eli käyttäjän itse syöttämiä hakusanoja enemmän sisältöjen tarjonnan pohjana. YouTuben suositukset painottavat personoinnin lisäksi videoita, jotka ovat olleet vetovoimaisia muille käyttäjille. Esimerkiksi syötettäessä haku YouTubeen vain noin 8 ensimmäistä hakutulosta liittyy varsinaiseen hakukenttään syötettyyn hakuun ja loput ovat suosituksia (”Muut katselivat myös näitä,” ”Sinulle,” ”Tähän liittyvistä hauista” ja ”Myös näitä haetaan”). Pahimmassa tapauksessa nämä liittyvät vain liepeellisesti tehtyyn hakuun ja hakutermeihin. Valtaosa YouTuben käyttäjien videovalinnoista (noin 70 %) tapahtuukin sivun oikeaan laitaan ilmestyvien suositusten, ei hakujen perusteella.²⁴ Netflixissä 80 prosenttia käyttäjien valinnoista perustuu suosituksiin ja 20 prosenttia hakuihin.²⁵ Koska haut Netflixissä koskevat enimmäkseen aineistoja, joita palvelussa ei ole, hakutuloksetkin ovat luonteeltaan sisältöperustaisia suosituksia hakuun läheisesti liittyvistä sisällöistä.

Suosituspainotteisuus haun pohjana toimii myös siinä, että Google- ja YouTube-haut antavat automaattisesti vinkkejä haun tarkentamiseksi muiden ihmisten tekemien hakujen perusteella. Käyttäjät, joilla on tarkkoja tiedontarpeita, saattavat kuitenkin helposti turhautua suosituspainotteiseen hakuun. On myös mahdollista, että käyttäjän oma tiedontarve tai haku muokkaantuu näiden muiden tekemien hakujen perusteella. Käyttäjän kontrollin kannalta tämä on

²³ Willson 2017.

²⁴ Southern 2021.

²⁵ Gomez-Uribe & Hunt 2015.

ongelmallista. Suositusperustainen haku on suhteellisen uusi asia, joka kaipaa enemmän kokeellista kriittistä tiedonhakututkimusta.

YouTuben, Instagramin ja X:n (aiemmin Twitter) tarjoamat keinot käyttäjälle muokata omia suosituksiaan ja uutissyötettään ovat tutkimusten mukaan näennäisiä. Algoritmien oikeudenmukaisuutta tutkittaessa on havaittu ilmiö, jota kutsutaan *suosituimmuusvinoumaksi* (*popularity bias*).²⁶ Periaatteessa suositusten pitäisi siis heijastaa johdonmukaisesti käyttäjien omia aiempia valintoja ja pitämisiä. Käyttäjäryhmät voidaan erotella suuntautumistensa osalta monipuolisia ja vaihtelevia sisältöä kuluttaviin, erityisintressejä omaaviin (*niche-focused*) ja hittisisältöihin keskittyviin.²⁷ Suositukset, joita käyttäjät saavat, ensisijaistavat hittisisällöt ja uutuustarjonnan. Niin sanotun pitkän hännän sisällöt eli pienempien erikoistuneiden yleisöjen suosimat ja harvinaisemmat sisällöt jäävät helpommin piiloon.

X:n ”trendaavat aiheet” tai ”uutiset” sekä Netflixin kategoriat ”suosittua alueellasi” tai ”mikä on uutta” eivät perustu käyttäjän omaan profiliin. Alustat toimivat näiden suositustyyppien osalta enemmän kuten perinteiset kaupalliset sisällöntuottajat ja -markkinoijat. Käyttäjille annetaan kuitenkin monilla alustoilla virheellinen illuusio kontrollista. Esimerkiksi YouTube tarjoaa suositusten yhteydessä toiminnot ”älä suosittele kanavaa” ja ”en ole kiinnostunut”. Nämä toiminnot eivät kuitenkaan todellisuudessa vaikuta suosituksiin lainkaan tai niillä on vaikutusta vain väliaikaisesti.²⁸ Käyttäjä ei todellisuudessa pysty sulkemaan suosituksistaan pois esimerkiksi Englannin kuninkaallisia tai Hollywood-julkkiksia koskevaa uutisointia.

Suositusalgoritmien ja käyttäjien kontrollin kannalta on keskeistä, voivatko käyttäjät sulkea ei-kiinnostavia aihealueita suositustensa ja uutissyötteensä ulkopuolelle. ”Älä suosittele kanavaa”-toiminto koskee vain yksittäistä kanavaa, mutta ei muuta saman-

26 Abdollahpouri ym. 2019.

27 Abdollahpouri ym. 2019.

28 de Luna 2022; Perry 2022; Ricks & McCrosky 2022.

tyyppistä sisältöä. ”En ole kiinnostunut” -toiminnolla ei laajan kokeellisen ja haastattelututkimuksen perusteella ole minkäänlaista vaikutusta suosituksiin.²⁹ Uutissyötettä muovaa käyttäjän aiemman toiminnan ohella yhtä paljon samanlainen kaupallinen logiikka kuin (ilta)päivälehtien uutisointia, eikä käyttäjällä ole mahdollisuuksia kontrolloida tai kritisoida tämän tyyppistä tiedontarjontaa.³⁰

Netflixistä ja Spotifystä poiketen YouTube ei ole luokittanut massiivista sisältöään näkyvästi genreihin tai aihealueisiin siten, että sisällöntuottajat saisivat tuetusti sijoitettua videonsa oikeisiin kategorioihin. Mikäli YouTube kehittäisi sisältöperustaista suositusta esimerkiksi jakamalla eri aihealueet omiin suppeampiin kanaviinsa YouTube sisällä ja pakottaisi sisällöntuottajat sijoittamaan tuotteensa näihin aiheperustaisiin kanaviin, mahdollisuus käyttäjien omiin valintoihin lisääntyisi. Alustan näkökulmasta luonnollisesti tärkeintä on kuitenkin löytää keinoja saada käyttäjät pysymään alustalla mahdollisimman tiiviisti mainontatulojen saamiseksi.

Suosituimmuus ja ajankohtaisuus suosituserusteina perustuvat luonnollisesti osaltaan käyttäjien (enemmistön) valintoihin. Suosio vaikuttaa osaltaan myös tarjonnan kehittymiseen. Enemmistön tekemät valinnat ja tuotteiden suosio vaikuttavat siihen, millaisia uusia sisältöjä esimerkiksi Netflix tuottaa tai millaisia kanavia YouTube nostaa näkyvästi esille. YouTubella on partneriohjelma, johon pääsy mahdollistaa alustan määrittelemien ehtojen puitteissa tulojen saamisen sisällöntuottajana.³¹ Sisällöntuottajan näkökulmasta mahdollisuus tulojen menetykseen (*demonetization*) saattaa motivoida välttelemään arkaluonteisia tai kiistanalaisia aiheita.

Käyttäjät pyrkivät myös kouluttamaan algoritmeja ymmärtämään heitä paremmin.³² He pyrkivät kontrolloimaan suosituksia ja saamaan niitä vastaamaan enemmän todellisia intressejään. Käyttäjät saattavat esimerkiksi hakea omien pysyvien intressiensä kannata

29 Ricks & McCrosky 2022.

30 Thorson ym. 2021.

31 Ks. <https://support.google.com/youtube/answer/72851?hl=en>

32 Gillespie 2014.

epätyypillisiä sisältöjä kirjautumatta sisälle palveluun tai luoda eri tilanteita ja mielenkiinnon kohteita varten eri profileja.

Tärkeää on kuitenkin huomata, että algoritmien noudattamat ohjeet ja säännöt perustuvat suurelta osin myös koneoppimiseen. Toisin sanoen käyttäjinä olemme osallisia algoritmien muovautumisessa. Tuotamme eksplisiittisesti dataa hakujen, ostosten, klikkauksen, pitämisten ja kommenttien muodossa. Eksplisiittisten reaktioiden lisäksi alustoille ja sivustoille kerääntyvät datamassat kertovat tarkemmin siitä, mikä on ollut erilaisten sisältöjen tavoittama katsojamäärä ja keitä olemme olleet maantieteellisesti, sukupuoleltamme ja iältämme.

Verkkopalvelut käyttävät evästeitä (*cookies*), joilla kerätään dataa käyttäjistä ja heidän toiminnastaan verkossa ja sivustolla. Euroopan unionin tietosuojalaki (GDPR) velvoittaa sivustot ilmoittamaan käyttäjille evästeistä sekä heitä hyväksymään tai hylkäämään ne.³³ Käytännössä tällä ei ole merkitystä muuten kuin käyttäjän informoinnin osalta.³⁴

Koneoppimista ja käyttäjien reaktioiden vaikutusta datakorpukseen on helpointa ymmärtää kuvantunnistusalgoritmien kehitymisessä. Esimerkiksi kuvienjakamispalvelu Flickr tarjoaa kuvia palveluun ladattaessa automaattisesti kuvailuja, jotka käyttäjä voi hylätä tai hyväksyä. Se ehdottaa, että kuvassa on esimerkiksi eläin, auto tai rakennus. Jos osuma on oikea, tämä opettaa algoritmia edelleen ja kuvantunnistusohjelmat tarkentuvat. Koneoppimisen ja palaute-luupin kautta algoritmit parantavat itseään: tekevät korrelointeja, suodattavat, kohdentavat ja oppivat ennustamaan ja ennakoimaan käyttäjien toimintaa.

Koska käyttäjien vuorovaikutus algoritmien kanssa vaikuttaa palaute-luuppina koneoppimisen kautta algoritmien kehitykseen, käyttäjät tavallaan saavat sellaisen infrastruktuurin kuin ansaitsevat. Jos siis enemmistö tarttuu suosituimpien sisältöjen suosituksiin, tämä

33 Bauer ym. 2022.

34 Eskens 2020.

osaltaan yhdenmukaistaa palauteluupin kautta kulutusta ja tarjontaa. Monet käyttäjät kehittävät kuitenkin kiertoteitä algoritmien logiikan ympäri, ja samalla heille kertyy tärkeää kokemustietoa algoritmien toimintaperiaatteista.

Algoritminen profilointi, seuranta ja yksityisyys

Käyttäjän kontrollin toinen puoli koskee yksityisyyttä. Kaikki käyttäjät eivät halua tulla alustojen tunnistamiksi ja profiloituiksi.³⁵ Yksityisyyden suojaaminen voi olla periaatteellinen kanta tai käyttäjä ei halua persoonaansa kohdistuvan vääränlaisia, virheellisiä tai häiritseviä alustojen luomia luokituksia.³⁶ Kuten edellä todettiin, käyttäjät dataa hyödynnetään esimerkiksi kohdennetuissa mainoksissa ja tätä tarkoitusta varten sitä jaetaan myös alustojen kesken.³⁷ Monet käyttäjät datan käsittelyn piirteet ovatkin laillisuuden rajamailla – tai pikemminkin lainsäädäntö raahaa teknologisen kehityksen perässä.³⁸

Data ei luonnollisestikaan luo profileja käyttäjistä ilman olemassa olevaa luokitusjärjestelmää.³⁹ Esimerkiksi Facebookilla on 20 000 datapistettä jokaisesta käyttäjästä.⁴⁰ Mikroprofilointi on luonteeltaan psykograafista eli se mittaa personallisuuden piirteitä, arvoja, mielipiteitä (uskonto ja poliittinen suuntautuminen) ja elämäntyyplejä. Tämä tuottaa hyvinkin hienojakoista tietoa jokaisesta yksittäisestä käyttäjästä. Laajaan vertailuaineistoon (niin sanottu *big data*) pohjautuen pienetkin asiat, kuten se, tykkäämmekö kukkakuvista tai kissavideoista tai välitämmekö eteenpäin sotauutisia, paljastavat lopulta meistä hyvinkin paljon.⁴¹

35 Burrell ym. 2019.

36 Burrell ym. 2019.

37 Büchi ym. 2021.

38 Baik 2022; Lomas 2022.

39 Kitchin 2014.

40 Tinker 2018.

41 Tinker 2018.

Periaatteessa Facebookin laajaan dataan ei saisi lainsäädännön mukaan yhdistää yksilöiviä henkilötietoja. Emme voi kuitenkaan varmasti tietää, hyödynnetäänkö psykograafista profilointia kohdennetun mainonnan ulkopuolella esimerkiksi käyttäjiin vaikuttamiseen. Kokeita tehdään ainakin tutkimustarkoituksessa siitä, millaisia tunteita erilaiset syötteet herättävät.⁴² Cambridge Analytica-skandaali osoitti esimerkiksi Facebookin olevan kokeilukenttä sen suhteen, voidaanko uutissyötteen tarjonnalla vaikuttaa emotionaalisen tilan ohella ihmisten mielipiteisiin ja poliittisiin suuntautumisiin.⁴³

Laaja-alaisempi taivuttelu tapahtuu Facebookissa näkyvästi eli rankaisemalla ja huomauttamalla ”väärän informaation” postittajia esimerkiksi varoituksin tai väliaikaisin ja lopullisin käyttökieltoin.⁴⁴ Faktantarkistukseen ja profiilien poistoon liittyvissä oikeudenkäynneissä Facebook on kuitenkin joutunut myöntämään, että alustan viralliselta vaikuttava faktantarkistus edustaa vain alustan omaa mielipidettä ja näkökulmaa. Twitter-kansioiden paljastukset toivat selkeästi esille, että todellisten sääntörikkomusten perusteella tehtyjen profiilien poistojen lisäksi rajoittamisella tähdätään poliittiseen tilanteeseen liittyvien narratiivien hallintaan.⁴⁵ Poliittisen tilanteen muuttuessa ”faktat” saattavat kääntyä pääläelleen.⁴⁶ Snowdenin vuoden 2013 paljastusten sekä Twitter-kansioiden paljastusten jälkeen tiedetään, että seuranta ja sensuuri tapahtuu valtioiden ja teknologisten alustojen yhteistoimintana.⁴⁷ Yksityisomisteisilla teknologiayrityksillä on luonnollisesti oikeus rajoittaa toimintaa alustoiltaan, kun taas valtiollisilla toimijoilla ei periaatteessa olisi laillista oikeutta takaporttien kautta tapahtuvaan laajaan henkilöseurantaan tai perustuslaillisen sananvapauden laaja-alaiseen rajoittamiseen.⁴⁸

42 Selinger & Hartzog 2016.

43 Baik 2022; Selinger & Hartzog 2016.

44 Gillespie 2022.

45 Davidson 2023.

46 Kaplan 2025.

47 Büchi ym. 2020; Davidson 2023; Gorwa ym. 2020.

48 Langvardt 2017.

Käyttäjien suhtautumista algoritmiseen sensuuriin tutkittaessa on kuitenkin havaittu, että ihmisillä on taipumus ajatella, että niin sanottu suuri yleisö eli muut ihmiset ovat alttiimpia manipulatiolle, informaatiovaikuttamiselle tai väärälle informaatiolle kuin he itse. Tätä kutsutaan *kolmannen persoonan teoriaksi*.⁴⁹ Ihmiset uskovat itse pystyvänsä arvioimaan sivustojen laatua ja tietoväittämien osuvuutta, mutta ajattelevat, että merkittävä osa muista ihmisistä ei pysty tähän.⁵⁰

Käyttäjien suhtautumista algoritmiseen profilointiin on tutkittu havainnoimalla käyttäjien reaktioita oman Facebook-profilinsa sisältämiin kohdennettua mainontaa varten luotuihin luokituksiin.⁵¹ Tutkimuksessa käyttäjiä pyydettiin tarkastelemaan ja kommentoimaan Facebookissa olevia kohtia ”Your interests” ja ”Your categories”. Nämä algoritmisesti muodostuvat mainontaprofilit ovat käyttäjille näkyvillä ja tarkastettavissa. Tosin käyttäjän täytyy osata edetä melko syvälle oman käyttäjäprofilinsa asetuksiin löytääkseen kohdat.

Yllättävän suuri osa (noin 40 %) vastaajista oli tietämätön siitä, että Facebook luokittelee heidän mielenkiinnon kohteitaan kohdennettua mainontaa ja uutissyötteen moderointia varten.⁵² Samanlaisia tuloksia on saatu muissakin tutkimuksissa.⁵³ Osa vastaajista järkyttyi ja huolestui huomattavasti, että osa heidän profilistaan oli muodostunut esimerkiksi yksityisten viestien ja Facebookin ulkopuolella suoritettujen verkkohakujen perusteella eli muulla tavoin kuin heidän omien Facebook-postaustensa perusteella.⁵⁴ Profilissa oli merkkejä koulutustasoa, ammattia, parisuhdetta ja perhettä koskevista tiedoista, joita käyttäjä ei ollut tietoisesti antanut palvelulle.

49 Davison 1983; McLeod ym. 2001; Riedl ym. 2022.

50 Dogruel ym. 2022; Rojas ym. 1996.

51 Büchi ym. 2021.

52 Büchi ym. 2021.

53 Hitlin & Rainie 2019; Powers 2017.

54 Büchi ym. 2021.

Jotkin kohdennetut mainokset näyttivät perustuvan yksityisiin puhelinkeskusteluihin.⁵⁵

Facebookin kohdennettujen mainosten osalta käyttäjän on mahdollisuus rastittaa pois kohdaltaan niitä intressejä ja kategorioita, joita hän ei pidä osuvina. Koska profilit muodostuvat algoritmisesti, eivätkä ne siten ole staattisia, käyttäjän omalla mainontaprofilinsa muokkaamisella ei ole pitkäaikaista vaikutusta.⁵⁶ Kaikki toiminta Facebookissa ja sen ulkopuolella eli tykkäykset, kommentit, kirjoitetut tekstit, hakukoneeseen kirjoitetut hakusanat, verkko-ostokset, jopa yksityinen puhe ja viestit kännykässä, vaikuttavat intressiprofilin muodostumiseen. Käyttäjälle näytetty mainontaa varten luotu karkea intressiprofiili ei ole sama asia kuin algoritmisesti muodostettu psykografinen mikroprofiili. Facebook siis esittää käyttäjälle itselleen päättelämänsä kategoriat ja intressit karkealla tasolla, kun tosiasiaa luokituskriteerejä ja -aspekteja on tuhansittain.

Käyttäjät eivät ole välttämättä tietoisia siitä, että Facebook ja muut palveluntarjoajat (YouTube, Google, Microsoft ja Amazon) jakavat dataa keskenään tai että Facebook-ystävien toiminta (tykkäykset, kommentit ja ystävien seuraamat sivut) vaikuttavat profilointiin.⁵⁷ Käyttäjakohtainen algoritmisen profilointi on alustat ylittävää eli eri alustojen dataa yhdistelevää.⁵⁸ Näin on epäilemättä myös profilointiarkkitehtuurin suhteen: alustojen datapisteiden eli ihmisiä ja sisältöjä koskevien luokitusten ja kategorisointien täytyy olla keskenään yhteensopivia, jotta dataa voidaan myydä ja vaihtaa eri osapuolten kesken.

Facebook-profilin kategorisointien osalta osalle käyttäjistä oli yllätys, miten tyhmiä ja epävakuuttavia profilit olivat. Maantieteellisen sijainnin, iän ja sukupuolen perusteella muodostetut profilit eivät useiden käyttäjien mielestä millään lailla vastanneet heidän to-

55 Büchi ym. 2021.

56 Thorson ym. 2021.

57 Büchi ym. 2020; Büchi ym. 2021.

58 Büchi ym. 2020.

dellisiä intressejään tai verkkokäyttäytymistään.⁵⁹ Näiden käyttäjien mielestä algoritmien personointikyvykyys ei ollut tavoittanut heidän todellista persoonaansa. Monet olivat odottaneet tekoälyn kehittyneen pidemmälle personoinnin suhteen. Demografinen profilointi vaikutti jäykältä ja stereotyyppiseltä.⁶⁰

Kohdennetun mainonnan osalta alustojen (Facebook, Instagram, X ja YouTube) myytävä ja markkinoima kohde ovat käyttäjät itse ja käyttäjädata. Profilointi tarkoittaa systemaattista ja tarkoituksellista ihmisiin liittyvää luokittelua.⁶¹ Automaattiset algoritmiset profiloinnit – niin harhaanjohtavia kuin ne persoonakohtaisesti saattavatkin olla – muovaavat omalla tavallaan yhteiskuntaa, elämyksiä ja elämismaaailmaa.⁶²

Bucher tutki haastattelemalla käyttäjien havaintoja ja kokemuksia Facebookin uutissyötealgoritmista.⁶³ Ärtymystä ja loukkaantumista aiheuttava seikka oli esimerkiksi se, että kohdennetut mainokset saattoivat puskea stereotyyppistä käsitystä esimerkiksi keski-ikäisten naisten huolenaiheista syöttämällä mainoksia painonpudotustuotteista ja ryppyvoiteista.⁶⁴ Tämä tuotti epämiellyttävän törmäämisen itseen ulkopuolisen näkökulmasta. Algoritmin ”silmät” voivat olla armottomat.

Sukupuolentunnistusalgoritmit tunnistavat käyttäjän sukupuolen, vaikka käyttäjä haluaisi salata sukupuolensa ja ikänsä ja tällä tavoin vaikeuttaa itseään koskevan datavarannon muodostumista. Tekstipohjaiset sukupuolentunnistusjärjestelmät on kehitetty analysoimalla sisältöjä ja tekstipiirteitä sähköposteista, blogeista, chatista ja sosiaalisesta mediasta.⁶⁵ Järjestelmät vertaavat automaattisesti käyttäjän tuottamia tekstejä korpusdataan, jossa ilmaisutavat,

59 Büchi ym. 2021.

60 Büchi ym. 2021.

61 Büchi ym. 2020.

62 Willson 2017.

63 Bucher 2017.

64 Bucher 2017.

65 Fosch-Villaronga ym. 2021.

tyylipiirteet ja kielenkäyttö on luokiteltu joko maskuliiniseksi tai feminiiniseksi.⁶⁶

Algoritmit ymmärretään usein neutraaleiksi ja suhteellisen objektiivisiksi,⁶⁷ mutta yhteiskunnalliset stereotypiat saattavat koneoppimisluupin kautta vahvistua ilman, että tämä on algoritmin kehittäjien alkuperäinen tavoite. Tutkimuksissa on havaittu, että sukupuolistereotypiat näkyvät esimerkiksi YouTube-videosuosituksissa.⁶⁸ Naisille suositellaan esimerkiksi meikkaus-, ruuanlaitto- ja käsityövideoita, vaikka he eivät olisi koskaan etsineet näitä sisältöjä YouTubea.⁶⁹ Miehillä saatetaan suositella enemmän esimerkiksi pelaamista, jalkapalloa, teknologiaa, autoja sekä huumori- ja sketsivideoita.⁷⁰

Päätoimisten sisällöntuottajien kannalta pysyttäytyminen turvallisten ja jossain määrin konventionaalisten esitystapojen ja aiheiden parissa on tärkeää siksi, että alustat rankaisevat melko herkästi tottelemattomuudesta vähentämällä videoiden näkyvyyttä.⁷¹ X, Instagram, YouTube ja Facebook käyttävät varjobannausta, jolla ne rajoittavat käyttösääntöjensä perusteella ei-toivotun tai rajatapaukseksi luokitellun sisällön näkyvyyttä.⁷² Varjobannauksella tarkoitetaan sitä, että käyttäjän postausta tai sisältöä ei näytetä kaikille niille, jotka ovat sisällön tilanneet.⁷³ Käyttäjä havaitsee varjobannauksen siitä, etteivät postaukset saavuta samoja katsontamääriä kuin aiemmin tai eivät lainkaan näytä tavoittavan kanavan tilaajia ja katsojia. Varjobannaukseen johtaa esimerkiksi rajoitettujen hashtagien eli aihetunnisteiden käyttö.⁷⁴ Julkaisualustojen kirjalliset yhteisönormit (*community guidelines*) ovat melko väljiä ja tul-

66 Fosch-Villaronga ym. 2021.

67 Gillespie 2014; Kitchin 2017.

68 Bishop 2018.

69 Bishop 2018.

70 Bishop 2018.

71 Gillespie 2022.

72 Cotter 2021.

73 Cotter 2021.

74 Cotter 2021.

kinnanvaraisia, eikä syitä varjobannaukselle tai profiilin sulkemiselle yleensä kerrota tarkasti.⁷⁵

Näkyvyyden rajoittamisen sattumanvaraisuutta lisää se, etteivät algoritmit pysty tulkitsemaan kulttuurista kontekstia. Algoritmi ei tunnista, onko postauksessa kyse roolipeliselostuksesta vaiko todellisesta tappamisuhkauksesta.⁷⁶ Algoritmi ei myöskään tiedä, onko postauksen alastonkuva kuva käyttäjästä vaiko taideteos tai levynkansi. Tämänkaltaisen arvaamattomuus luonnollisesti lisää painetta itsesensuuriin tai pysyttelyyn konventionaalisissa sisältöalueissa.⁷⁷ Käyttäjät voivat kuitenkin myös etsiä kiertoteitä algoritmisen sensuurin ohitse esimerkiksi kirjoittamalla tai lausumalla väärin sanoja, jotka ovat erityisessä seurannassa.

Algoritmilukutaito

Joidenkin tutkimusten mukaan käyttäjät eivät välttämättä ole kovin tietoisia siitä, kuinka algoritmit suodattavat uutissyötteitä esimerkiksi Facebookissa.⁷⁸ Algoritmitietoisuudesta tai -lukutaidosta on jopa puhuttu toisen asteen digitaalisena kuiluna, koska algoritmien vaikutus informaation ja sisältöjen saatavuuteen on niin suuri.⁷⁹ Algoritmilukutaitoon kuuluu monia osa-alueita, kuten tietoisuus itseä koskevan tiedon keruusta, hakukoneiden toimintaperiaatteista sekä suodatus- ja suositusperusteista eli siitä, millä kriteereillä sisältöjä valikoidaan tarjottavaksi eri palveluissa. Kyky vaikuttaa esimerkiksi uutissyötteisiin tietoisesti ja kriittisesti ja etsiä vaihtoehtoja standardituloksille tai -tarjonnalle on tärkeää. Näin on varsinkin haluttaessa mahdollisimman laaja perspektiivi jostakin aiheesta käytyihin keskusteluihin esimerkiksi tutkimustarkoituksessa tai mikäli ajatus ja

75 Gillespie 2022.

76 Burrell ym. 2019.

77 Gillespie 2022.

78 Powers 2017.

79 Gran ym. 2021.

tavoite tietyn teknologisen alustan käytöstä eroaa siihen sisäänrakennetuista oletuksista käyttäjän toiminnasta.

Norjalaisessa kyselytutkimuksessa⁸⁰ testattiin alustariippumattomaa yleistä algoritmitietoisuutta. Kyselyn tulosten mukaan 41 prosenttia norjalaisista katsoi olevansa tietämätön algoritmien toiminnasta, 47 prosenttia jossain määrin tietoinen ja 10 prosenttia erittäin tietoinen. Ei-asiantunteviksi itsensä lukevat suhtautuivat yleensä positiivisesti suosituksiin sekä positiivisesti tai neutraalisti personoituun mainontaan. Algoritmeja tuntevien ryhmä näki suositukset ja uutissuodatuksen sekä positiivisena että negatiivisena. Tämä ryhmä oli eniten huolestunut algoritmisesta sensuurista sekä käyttäjien seurannasta ja kontrolloinnista. Algoritmitietoisuuden ja -tuntemuksen voidaan ajatella vaikuttavan yhteiskunnalliseen oikeudenmukaisuuteen ja yleisiin demokraattisen osallistumisen mahdollisuuksiin.

Profilointi on tuttu termi varsinkin rikossarjoista, eikä vertaus ole sikäli kovin kaukaa haettu, että yksi alkuperäinen tarkoitus käyttäjädatan massiivisessa keräämisessä on ollut tunnistaa ja seurata mahdollisia terrorismiuhkia. Yhä tarkempien datankeruumenetelmien, esimerkiksi kasvojen- ja puheentunnistuksen, taustalla ei koskaan ole vain halu kehittää teknologiaa sen itsensä vuoksi.⁸¹ Tarkempi kohteiden tunnistaminen, oli kyse sitten kuvissa olevissa esineistä tai ihmisistä, fyysisessä tai digitaalisessa ympäristössä, on filosofisessa mielessä objektivismin harhan läpikäynti.⁸² Kuten edellä todettiin, algoritmit eivät kykene tulkitsemaan esimerkiksi kuvien kontekstia. Ilman tietoa kuvan kontekstista ja kulttuurisesta ympäristöstä tunnistaminen voi tapahtua vain pinnallisella tasolla.

Suchman huomauttaa, ettei teknologian kehittäjien pyrkimys kohteiden yhä tarkempaan tunnistamiseen, määrittelyyn ja luokiteluun ole missään tapauksessa vähenemässä.⁸³ Teknologioiden ke-

80 Gran ym. 2021.

81 Suchman 2022.

82 Suchman 2022.

83 Suchman 2022, 11.

hittäjillä ei ole mielessään yhtenä vaihtoehtona ajatusta, että terrorismin uhan tai sodankäynnin tarpeet yhtenä keskeisenä motivaattorina tunnistamisteknologioiden kehittämisessä jonakin päivänä loppuisivat. Monesti ajatellaan, että profiloinnin kehittämisen tarpeet kumpuavat ennen muuta mainonnan ja myynnin edistämisen tarpeista, mutta, kuten internet kokonaisuudessaan, erilaiset seurannan ja profiloinnin teknologiat ovat alun perin syntyneet sotilaallisteknologisen kompleksin rahoittamana ja alullepanemana.⁸⁴

Tutkimusten mukaan ihmisten suhtautuminen seurantaan edustaa tyypillisesti valvontarealismia tai valvontauupumusta.⁸⁵ Ihmiset ovat tietoisia seurannasta, mutta enimmäkseen he katsovat, ettei systemaattiselta oman datansa keruulta voi mitenkään välttää tai että välttäminen vaatisi erityisasiantuntemusta ja suuria ponnistuksia. Ajatus seurannasta on epämiellyttävä, mutta toisaalta seuranta on normalisoitunut.⁸⁶ Yksityisyyteen kohdistuvat uhat konkretisoituvat silloin, kun uutisissa kerrotaan käyttäjätietojen tai salasanojen vuodoista. Useimmat käyttäjät uskovat, että heidän todellinen minänsä pysyy kuitenkin suojattuna ja muuttumattomana datankeruun ylilyönneistä huolimatta.⁸⁷

Algoritmien rajallinen kyky tulkita kontekstia koskee myös aikaa: erityisesti Facebookissa mutta myös muilla alustoilla algoritmeilla ei ole aina kykyä huomioida oleellisia käännekohtia tai siirtymiä käyttäjän intresseissä. Facebookin ja muiden algoritmien kyky tuoda esille käyttäjän menneisyyteen kuuluvia asioita omasta aloitteestaan saattaa aiheuttaa epämiellyttäviä tunteita.⁸⁸ Esimerkiksi Facebook tarjoaa käyttäjälleen automaattisesti nostoja ja postitusehdotuksia kuten menneisyyden tapahtumia (Muistot -osio, Tänä päivänä -sivu, Kohokohdat). Algoritmi ei pysty tulkitsemaan sitä, ovatko tapahtumat, kontaktit tai vuorovaikutus positiivisesti tai ne-

84 Suchman 2022.

85 Dencik & Cable 2017.

86 Dencik & Cable 2017.

87 Lupton 2020.

88 Bucher 2017; Jacobsen 2022.

gatiivisesti sävyttyneitä. Ne eivät ymmärrä, millainen informaatio on todellisuudessa ajankohtaista tai osuvaa käyttäjän omassa kontekstissa.⁸⁹

Facebookin algoritmia muutettiin oleellisesti vuonna 2018.⁹⁰ Sitä ennen uutissyöte järjestyi kronologisesti niin, että uusimmat postaukset olivat automaattisesti syötteessä ylinnä. Muutoksen jälkeen Facebook ensisijaisti lähimpien ystävien ja perheen postaukset haluten palata palvelun alkuperäiseen ystäväkirjaluonteeseen.⁹¹ Facebook halusi rohkaista nimenomaan ”autenttista” henkilökohtaista vuorovaikutusta erilaisten yhteiskunnallisten asioiden ja keskustelun tai harrastusintressien seuraamisen sijaan.⁹² Facebook näyttää uutissyötteessä postauksia ensisijaisesti vain niiltä henkilöiltä, joiden kanssa käyttäjä on eniten vuorovaikutuksessa. Tästä johtuen käyttäjät ovat kokeneet uutissyötteen muuttuneen suppeaksi ja toisteiseksi.⁹³ Sama koskee Instagramia.⁹⁴ Käyttäjät voivat pyrkiä laajentamaan uutissyötettään käymällä klikkaamassa laajemmalla skaalalla profileja ja Facebook-sivuja, mutta tämä vaatii tietoista vaivannäköä.⁹⁵

Lopuksi

Algoritmien vaikutusta maailmamme ja arkielämämme muodostumiseen ei voi aliarvioida. Enimmäkseen mahdollisuutemme tietää ja ymmärtää sitä, missä määrin kohtaamamme informaatio, tieto ja digitaalinen todellisuus on algoritmien puitteistamaa ja rajaamaa, on hyvinkin puutteellinen. Algoritmien tarkka koostumus on liikesalaisuus, ja sikäli käyttäjien omat havainnot ja huomiot eivät ole mitenkään toissijaisia algoritmitutkimuksessa. Lisäksi algoritmien

89 Jacobsen 2022.

90 Mosseri 2018.

91 Mosseri 2018.

92 Lima 2021; Mosseri 2018.

93 Dogruel ym. 2022.

94 Dogruel ym. 2022.

95 Bucher 2017.

toiminnan tiedostamisesta ja ymmärtämisestä on tullut uusi digitaalinen kuilu niiden välille, jotka osaavat hakea kiertoteitä esimerkiksi algoritmisen sensuurin ohitse, ja niiden välille, jotka eivät ole tietoisia esimerkiksi hakukoneiden lähdepreferensseistä ja näytettyjen sisältöjen valikointi- ja rankkaamisperusteista.

Hakutulosten ja uutissyötteiden suodattajina algoritmit puitteistavat sosiaaliset kysymykset, puheenaiheet ja lähestymistavat joko normaaleiksi tai ei-normaaleiksi sekä legitiimeiksi tai ei-legitiimeiksi.⁹⁶ Ne esittelevät joko implisiittisesti tai eksplisiittisesti, kuinka yhteiskunnan jäsenen kuulua asennoitua ja käyttäytyä, mitä yleiset normit edellyttävät. Perinnemedian tavoin algoritmien suodatus muovaa sitä, mitä ihmiset tietävät sekä ketä he tuntevat. Perinnemedian, varsinkin julkisesti tuetun median, on ajateltu informoivan ihmisiä siitä, mitä heidän on tärkeää tietää. Algoritmit yrittävät myös arvailla, mitä ihmiset haluaisivat tietää. Digitaalinen minä eli algoritmien subjekti ei kuitenkaan monestakaan syystä ole todellisen käyttäjän kaksonen.

Yhteiskunnan tarve riippumattomalle algoritmitutkimukselle on suuri. Tarvitaan enemmän ymmärrystä siitä, missä määrin kertyvät datamassat kuvaavat todellisten ihmisten todellisia tarpeita, käyttäytymistä ja asenteita. Datapisteiden määrittely on tässä suhteessa ydinasia. Ihmisten ja intressien luokitukset eivät välttämättä vastaa ihmisten moninaisia suuntautumis- ja asennoitumisia. On hyvin mahdollista, että dataa kerätään asioista, jotka ovat toissijaisia, jotka eivät ole helposti luokiteltavissa ja jotka vaihtelevat tilanteiden ja kontekstien mukaan enemmän kuin mitä luokitusjärjestelmät voivat tavoittaa. Datapisteiden määrittely voi tuottaa itseään toistavan kehän, joka saattaa sulkea ulkopuolelleen vähemmän tunnettuja yhteisöjä, tiedon, harrastuksen ja taiteen aloja sekä ajatuksia ja kulttuureita. Luokitus on algoritmien ydintä.

Kuinka algoritmeja sitten voidaan tutkia? Tässä artikkelissa on esitelty esimerkkejä tutkimuksista, kuten käyttäjien tyytyväisyy-

96 Gillespie 2022; Lundahl 2022.

destä suosituksiin eri alustoilla (kuten YouTube ja Netflix). Käyttäjien näkemykset siitä, kuinka he haluaisivat suosituksia kehitettävän, ovat hyödyllisiä. Erityisen kiinnostavia ovat ne käyttäjät, jotka yrittävät löytää kiertoteitä ja huijaamisstrategioita saadakseen itselleen paremmin sopivia suosituksia ja sulkeakseen pois ei-sopivia suosituksia. Kokeellisen tutkimuksen osalta eri käyttäjien samoilla hauilla saamien hakutulosten ja suositusten vertailu kertoo persoonoinnin laajuudesta ja algoritmien mahdollisista puolueellisuuksista ja vinoumista esimerkiksi sukupuolen ja iän perusteella.

Algoritmien taustalla oleva datapisteiden ja käyttäjien tuottamien sisältöjen luokitus ja kuvailu ovat keskeisiä aiheita siitäkin huolimatta, ettei tuloksia hyödynnettäisi kaupallisten alustojen kehittämiseen. Myös julkiset palvelut kuten kirjastot voivat kehittää digitaalisia palvelujaan suosituserustaisten alustojen parhaita piirteitä hyödyntäen. Käyttäjien näkemykset autonomiastaan sekä ajatukset yksityisyydensuojasta ja seurannasta ovat luonnollisesti tärkeitä tutkimusaiheita. Empiiriset ja kokeelliset tutkimukset algoritmista sensuurista sekä aiheiden että käyttäjien osalta (kuten käyttäjien varjotilien poisto, väliaikaiset käyttökiellot tai tilien poisto) ovat erityisen tärkeitä. Jatkuva ristiriita perustuslakiin kuuluvan sananvapauden sekä algoritmisen ja suoran sensuurin välillä on keskeinen tutkimuskohde.

Lähteet

- Abdollahpouri, Himan & Masod Mansoury & Robin Burke & Bamshad Mobasher (2019) *The unfairness of popularity bias in recommendation*. arXiv preprint. <https://arxiv.org/pdf/1907.13286.pdf>
- Baik, Jeeyun (2022) Data privacy and political distrust: Corporate ‘pro liars,’ ‘gridlocked Congress,’ and the Twitter issue public around the US privacy legislation. *Information, communication & society* 25:9, 1211–1228. <https://doi.org/10.1080/1369118X.2020.1850839>
- Bauer, Paul C. & Frederic Gerdon & Florian Keusch & Frauke Kreuter & David Vannette (2022) Did the GDPR increase trust in data collectors? Evidence from observational and experimental data. *Information, communication & society* 25:14, 2101–2121. <https://doi.org/10.1080/1369118X.2021.1927138>
- Bishop, Sophie (2018) Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm. *Convergence* 24:1, 69–84. <https://doi.org/10.1177/1354856517736978>
- Brownstein, Barry (2019) Google is burying alternative health sites to protect people from “dangerous” medical advice. <https://intellectualexit.org/2019/08/google-is-burying-alternative-health-sites-to-protect-people-from-dangerous-medical-advice/>
- Bucher, Taina (2017) The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, communication & society* 20:1, 30–44. <https://doi.org/10.1080/1369118X.2016.1154086>
- Burrell, Jenna & Zoe Kahn & Anne Jonas & Daniel Griffin (2019) When users control the algorithms: Values expressed in practices on Twitter. *Proceedings of the ACM on human-computer interaction* 3(CSCW), 1–20. <https://doi.org/10.1145/3359240>
- Büchi, Moritz & Eduard Fosch-Villaronga & Christoph Lutz & Aurelia Tamò-Larrieux & Struthi Velidi & Salome Viljoen (2020) The chilling effects of algorithmic profiling: Mapping the issues. *Computer law & security review* 36, 105367. <https://doi.org/10.1016/j.clsr.2019.105367>
- Büchi, Moritz & Eduard Fosch-Villaronga & Christoph Lutz & Aurelia Tamò-Larrieux & Struthi Velidi (2021) Making sense of algorithmic profiling: User perceptions on Facebook. *Information, communication & society* 26:4, 1–17. <https://doi.org/10.1080/1369118X.2021.1989011>
- Cotter, Kelley (2021) “Shadowbanning is not a thing”: Black box gaslighting and the power to independently know and credibly critique algorithms. *Information, communication & society* 26:6, 1–18. <https://doi.org/10.1080/1369118X.2021.1994624>
- Davidson, John Daniel (2023) Twitter files reveal an existential threat. *Imprimis* 52:1. <https://imprimis.hillsdale.edu/the-twitter-files-reveal-an-existential-threat/>
- Davison, W. Phillips (1983) The third-person effect in communication. *Public opinion quarterly* 47:1, 1–15. <https://doi.org/10.1086/268763>
- De Luna, Elizabeth (2022) Sorry ‘Little Mermaid’ haters, YouTube’s dislike button doesn’t really do anything. <https://mashable.com/article/youtube-dislike-button-mozilla>

- Dean, Brian (2022) We analyzed 4 million Google search results: Here's what we learned about organic click-through rate. <https://backlinko.com/google-ctr-stats>
- Dencik, Lina & Jonathan Cable (2017) The advent of surveillance realism: Public opinion and activist responses to the Snowden leaks. *International journal of communication* 11, 763–781. <http://ijoc.org/index.php/ijoc/article/view/5524/1939>
- Dershowitz, Jessica (2024) The story behind Netflix's secret category codes. <https://www.netflix.com/tudum/articles/netflix-secret-codes-guide>
- Dogruel, Leyla & Dominique Facciorusso & Birgit Stark (2022) 'I'm still the master of the machine: Internet users' awareness of algorithmic decision-making and their perception of its effect on their autonomy. *Information, communication & society* 25:9, 1311–1332. <https://doi.org/10.1080/1366918X.2020.1863999>
- Dutta, Biswanath & Jyotima Patel (2022) Algorithm metadata vocabulary: A representational model and metadata vocabulary for describing and maintaining algorithms. *Journal of information science* 50:5, 1223–1241. <https://doi.org/10.1177/01655515221116557>
- Epstein, Robert (2016) The new censorship. *U.S. news*, 22.6.2016. <https://www.usnews.com/opinion/articles/2016-06-22/google-is-the-worlds-biggest-censor-and-its-power-must-be-regulated>
- Esken, Sarah (2020) The personal information sphere: An integral approach to privacy and related information and communication rights. *Journal of the association for information science and technology* 71:9, 1116–1128. <https://doi.org/10.1002/asi.24354>
- Fosch-Villaronga, Eduard & Adam Poulsen & Roger Andre Søråa & Bart H. M. Custers (2021) A little bird told me your gender: Gender inferences in social media. *Information Processing & Management* 58:3, 102541. <https://doi.org/10.1016/j.ipm.2021.102541>
- Gillespie, Tarleton (2014) The relevance of algorithms. Teoksessa Tarleton Gillespie & Pablo Boczkowski & Kirsten Foot (toim.) *Media technologies: Essays on communication, materiality, and society*. Cambridge, MA: MIT Press. https://intgovforum.org/multilingual/sites/default/files/webform/the_relevance_of_algorithms_-_tarleton_gillespie.pdf
- Gillespie, Tarleton (2022) Do not recommend? Reduction as a form of content moderation. *Social media + society* 8:3, 1–13. <https://doi.org/10.1177/20563051221117552>
- Goldman, Erik (2019) *How Google is controlling the flow of healthcare information*. <https://holisticprimarycare.net/topics/news-policy-a-economics/how-google-is-reshaping-the-flow-of-healthcare-information/>
- Gomez-Urbe, Carlos A. & Neil Hunt (2015) The Netflix recommender system: Algorithms, business value, and innovation. *ACM transactions on management information systems (TMIS)*, 6:4, 1–19. <https://doi.org/10.1145/2843948>
- Google (2022) General guidelines. <https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf>

- Gorwa, Robert & Reuben Binns & Christian Katzenbach (2020) Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big data & society* 7:1. <https://doi.org/10.1177/205395>
- Gran, Anne-Britt & Peter Booth & Taina Bucher (2021) To be or not to be algorithm aware: a question of a new digital divide? *Information, communication & society* 24:12, 1779–1796. <https://doi.org/10.1080/1369118X.2020.1736124>
- Grind, Kirsten & Sam Schechner & Robert McMillan & John West (2019) How Google interferes with its search algorithms and changes your results. *The Wall Street Journal*, 15. <https://www.wsj.com/articles/how-google-interferes-with-its-search-algorithms-and-changes-your-results-11573823753>
- Hitlin, Paul & Lee Rainie (2019) Facebook algorithms and personal data. *Pew Research Center* 1/2019 <https://s3.amazonaws.com/media.mediapost.com/uploads/FacebookAlgorithmsAndPersonalData.pdf>
- Hu, Charlotte (2021) Why Spotify's music recommendations always seem so spot on. <https://www.popsoci.com/technology/spotify-audio-recommendation-research/>
- Jacobsen, Benjamin N. (2022) Algorithms and the narration of past selves. *Information, communication & society* 25:8, 1082–1097. <https://doi.org/10.1080/1369118X.2020.1834603>
- Jones, Morgan (2022) Spotify's algorithm explained <https://bippermedia.com/spotify-s-algorithm-explained/>
- Kaplan, Joel (2025) More speech and fewer mistakes. <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>
- Kitchin, Rob (2014) Big Data, new epistemologies and paradigm shifts. *Big data & society* 1:1. <https://doi.org/10.1177/2053951714528481>
- Kitchin, Rob (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20:1, 14–29. <https://doi.org/10.1080/1369118X.2016.1154087>
- Langvardt, Kyle (2017) Regulating online content moderation. *Georgetown law journal* 106:5. <https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2018/07/Regulating-Online-Content-Moderation.pdf>
- Lekander, Alex (2025) 10 best private search engines in 2025 (Search anonymously). <https://restoreprivacy.com/private-search-engine/>
- Lima, Christiano (2021) Facebook to pilot new ways to reduce political content. *Politico*, 2.10.2021. <https://www.politico.com/news/2021/02/10/facebook-political-content-468253>
- Lomas, Natasha (2022) Unsealed docs in Facebook privacy suit offer glimpse of missing app audit. *TechCrunch*, 16.9.2022. <https://techcrunch.com/2022/09/16/unsealed-docs-in-facebook-privacy-suit-offer-glimpse-of-missing-app-audit/>
- Lundahl, Outi (2022) Algorithmic meta-capital: Bourdieusian analysis of social power through algorithms in media consumption. *Information, communication & society* 25:10, 1440–1455. <https://doi.org/10.1080/1369118X.2020.1864006>

- Lupton, Deborah (2020) Thinking with care about personal data profiling: A more-than-human approach. *International journal of communication* 14:19. <https://ijoc.org/index.php/ijoc/article/viewFile/13540/3114>
- MacLean, Jayson (2017) Google blacklists pseudoscience website Natural News. *Cantech letter*, 14.2.2017. <https://www.cantechletter.com/2017/02/google-blacklists-pseudoscience-website-natural-news/#>
- McLeod, Douglas M. & Benjamin H. Detenber & William P. Eveland, Jr. (2001) Behind the third-person effect: Differentiating perceptual processes for self and other. *Journal of communication* 51:4, 678–695. <https://doi.org/10.1111/j.1460-2466.2001.tb02902.x>
- Mosseri, Adam (2018) Bringing people closer together. <https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/>
- Perry, Alex (2022) Twitter's 'see less often' button is the ultimate internet placebo. <https://mashable.com/article/twitter-facebook-see-less-often-button>
- Powers, Elia (2017) My news feed is filtered? *Digital journalism* 5:10, 1315–1335. <https://doi.org/10.1080/21670811.2017.1286943>
- Rasmussen, Lukas (2021) Google rankings explained – How does Google decide which websites to rank when searching? <https://morningscore.io/how-does-google-rank-websites/>
- Ricks, Becca & McCrosky, Jesse (2022) Does this button work? Investigating YouTube's ineffective user controls. <https://foundation.mozilla.org/en/research/library/user-controls/report/>
- Riedl, Martin J. & Kelsey N. Whipple & Ryan Wallace (2022) Antecedents of support for social media content moderation and platform regulation: The role of presumed effects on self and others. *Information, communication & society* 25:11, 1632–1649. <https://doi.org/10.1080/1369118X.2021.1874040>
- Rojas, Hernando & Dhavan V. Shah & Ronald J. Faber (1996) For the good of others: Censorship and the third-person effect. *International journal of public opinion research* 8:2, 163–186. <https://doi.org/10.1093/ijpor/8.2.163>
- Selinger, Evan & Woodrow Hartzog (2016) Facebook's emotional contagion study and the ethical problem of co-opted identity in mediated environments where users lack control. *Research ethics* 12:1, 35–43. <https://doi.org/10.1177/1747016115579531>
- Shir-Raz, Yaffa & Ety Elisha & Brian Martin & Natti Ronel & Josh Guetzkow (2022) Censorship and suppression of Covid-19 heterodoxy: Tactics and counter-tactics. *Minerva* 61:3, 407–433. <https://doi.org/10.1007/s11024-022-09479-4>
- Southern, Matt G. (2021) 20 confirmed facts about YouTube's algorithm. *Search engine journal* 21.4.2021. <https://www.searchenginejournal.com/youtube-algorithm-facts/403984/#close>
- Suchman, Lucy (2022) Imaginaries of omniscience: Automating intelligence in the US Department of Defense. *Social studies of science* 53:5, 761–786. <https://doi.org/10.1177/03063127221104938>
- Thorson, Kjerstin & Kelley Cotter & Mel Medeiros & Chankyung Pak (2021) Algorithmic inference, political interest, and exposure to news and politics on Facebook.

3.3 Algoritmit ja tiedon löydettävyyys

- Information, communication & society* 24:2, 183–200. <https://doi.org/10.1080/1369118X.2019.1642934>
- Tinker, Ben (2018) How Facebook ‘likes’ predict race, religion and sexual orientation. *CNN*, 11.4.2018. <https://edition.cnn.com/2018/04/10/health/facebook-likes-psychographics/index.html>
- Tung, Liam (2019) Over half of Google searches don’t produce a single click: Is that good or bad? <https://www.zdnet.com/article/over-half-of-google-searches-dont-produce-a-single-click-is-that-good-or-bad/>
- Willson, Michele (2017) Algorithms (and the) everyday. *Information, communication & society* 20:1, 137–150. <https://doi.org/10.1080/1369118X.2016.1200645>

Informaatio yhteisöissä

4.1 Tieteellinen julkaiseminen – mitä, miksi ja miten?

Elina Late

- Tieteellisissä julkaisuissa raportoidaan alkuperäisiä tutkimustuloksia, jotka tyypillisimmin julkaistaan artikkeleina tieteellisissä lehdissä.
- Tieteenaloilla on omat vakiintuneet julkaisukäytäntönsä, jotka eroavat toisistaan.
- Vertaisarvointi on laadunvarmistusmekanismi, joka kuuluu tieteellisen kirjallisuuden julkaisuprosessiin.
- Open Access (OA) tarkoittaa tieteellisen tiedon avointa, vapaata saatavuutta. Avoimesti saatavilla oleva julkaisu on kenen tahansa luettavissa ilman tilausmaksua.
- Tieteellisiä julkaisuja tutkitaan sekä tutkimuksen että käytännön tarpeisiin, kuten tutkimuksen ja tutkijoiden arviointia varten.
- Informetrisillä menetelmillä tarkoitetaan määrällistä tutkimusta, jossa tutkimusaineistona hyödynnetään tietokannoista koottuja julkaisujen bibliografisia- ja viittaustietoja.

Tutkimus tuottaa jatkuvasti ja kiihtyvällä tahdilla uutta tietoa meitä ympäröivästä luonnosta, ihmisistä, teknologiasta, yhteiskunnasta ja kulttuurista. Keskeinen osa tutkijan työtä on välittää tuottamaansa tietoa eteenpäin, sillä ilman tulosten julkaisemista tavalla tai toisella niillä ei ole merkitystä kuin tekijälleen itselleen¹. Vaikka tutkimuksen tuloksena toisinaan syntyy uusia keksintöjä ja tuotteita, kuten teknologiaa tai vaikkapa lääkkeitä, tutkimuksessa tuotettu tieto jul-

1 Becher & Trowler 2001.

kaistaan teksteinä esimerkiksi artikkeleina, kirjoina tai raportteina. Julkaistua tietoa voidaan arvioida ja käyttää uuden tiedon pohjana sekä rakennuspalikoina. Julkaisemalla tuloksensa tutkija ottaa myös kunnian ja vastuun tekemästään työstä. Julkaisut ovatkin keskeinen osa tutkijan ansioluetteloa, ja niitä käytetään tutkijoiden ja tutkimuksen arvioinnin pohjana.

Tutkijat julkaisevat tuloksiaan monin eri tavoin ja erilaisille yleisöille. Tutkimustuloksista kirjoitetaan muun muassa sanoma- ja aikakauslehdissä, ja yhä useammin tutkijoita ja heidän työtään voi seurata myös sosiaalisessa mediassa. Alkuperäiset tutkimustulokset ja tutkimuksessa käytetyt menetelmät julkaistaan kuitenkin tieteellisillä foorumeilla, kuten tieteellisissä lehdissä, kirjoissa ja konferensseissa. Tieteelliset julkaisut ovat erikoistuneet tutkimustulosten julkaisemiseen, ja niillä on vakiintunut julkaisuprosessi, jossa pyritään varmistamaan julkaisujen laatu. Tulosten julkaisemisen jälkeen tutkijat odottavat tai ainakin toivovat huomiota ja tunnustusta tiedeyhteisöltä tai muulta lukijakunnalta. Tämä tapahtuu perinteisesti viittaamalla uusissa julkaisuissa jo aiemmin julkaistuun. Julkaisut ja niiden viittaukset muodostavatkin julkisen vuoropuhelun tutkijoiden välillä.

Tieteelliset julkaisut ovat eräänlainen ikkuna tutkimukseen. Niitä tarkastelemalla voidaan tutkia esimerkiksi tutkimusaiheiden ja uusien tutkimusalojen syntymistä, tutkijoiden tai julkaisujen vaikuttavuutta, tutkijoiden muodostamia verkostoja ja kansainvälisyyttä sekä eri tieteenalojen julkaisu- ja tiedontuotannon käytäntöjä. Informaatiotutkimuksen alalla onkin pitkä perinne tieteellisen julkaisemisen tutkimuksessa. Julkaisutoimintaa tarkastellaan myös käytännön tarpeisiin, joista tutkimuksen ja tutkijoiden arviointi ovat merkittävimpiä. Julkaisujen määrää ja niiden saamia viittauksia analysoidaan yliopistojen arvioinneissa tai vaikkapa tutkijoiden rekrytoinnissa ja tutkimusrahoituksen jaossa. Julkaisuja hyödynnetään myös osana niin sanottujen yliopistorankingien muodostamista, joissa yliopistoja listataan niiden merkittävyyden, näkyvyyden tai

menestyksen perusteella. Yliopistot seuraavat näitä listoja tarkkaan, sillä ne antavat yliopistoille näkyvyyttä ja mainetta.

Tutkimuksen mittaaminen ja arviointi julkaisujen perusteella ei kuitenkaan ole aivan yksinkertaista. Julkaisut eivät ensinnäkään kerro kaikkea tutkijan työstä, ja toisaalta eri tieteenaloilla on erilaiset tavat viestiä tutkimuksestaan. Arviointi herättää aina myös vastareaktion, mikä saattaa kannustaa tutkijoita ja julkaisijoita esimerkiksi tuottamaan suuria määriä julkaisuja laadun kustannuksella. Arviointi lisää myös kilpailua ja saattaa aiheuttaa tutkijoille painetta. Tunnettu sanonta *publish or perish* (julkaise tai tuhoudu) viittaa tutkijoiden paineeseen julkaista tutkimustuloksiaan edetäkseen urallaan².

Digitalisoituminen 2000-luvun aikana on muuttanut tieteellistä julkaisutoimintaa. Kun vielä 20 vuotta sitten tieto oli pääosin saatavilla tieteellisten kirjastojen ja painettujen julkaisujen kautta, nyt valtaosaan tiedosta on pääsy omalta tietokoneelta ja yhä useammin maksutta³. Digitalisaatio on mahdollistanut tiedon laajemman saatavuuden, ja yksi tärkeimmistä tämän hetken keskusteluista tutkimuksessa liittyykin tieteen avoimuuteen.⁴ Jo vuosikymmeniä on puhuttu informaatiotulvasta, ja tiedon määrän kasvu koettelee myös tutkijoita. Kaikkea omalta tieteenalalta tuotettua tietoa on mahdotonta seurata. Digitalisaatio on avannut tieteellisen julkaisemisen tutkimukseen myös uusia mahdollisuuksia, ja esimerkiksi laajoja tilastoaineistoja on tarjolla enemmän kuin koskaan. Tähän liittyy myös halu seurata ja mitata tutkimuksen tuottavuutta.

Tässä luvussa pyrin antamaan yleiskuvan tieteellisestä julkaisemisesta ja siihen liittyvistä ajankohtaisista keskustelunaiheista ja tutkimusmenetelmistä. Tarkastelen pääasiassa niin sanottua formaalia viestintää,⁵ jolla tarkoitetaan julkaistua kirjallisuutta. On kuitenkin tärkeää pitää mielessä, että tieteellinen julkaiseminen on vain osa

2 Karvonen ym. 2014; Van Dalen & Henkens 2012.

3 Borgman 2007; Johnson ym. 2018.

4 Esim. Tennant ym. 2019.

5 Meadows 1974

laajempaa tieteellistä viestintää ja tutkimustuloksia jaetaan monin tavoin, joista sosiaalinen media, erilaiset keskustelut ja opetus ovat hyviä esimerkkejä.

Mitä on tieteellinen julkaiseminen?

Julkaisemisella on merkittävä rooli tieteessä, ja sitä on kuvattu tutkimuksen ytimeksi, lahjojen antamiseksi tai myymiseksi ja ostamiseksi.⁶ Tutkimustuloksia viestitään erilaisille yleisöille, joista esimerkkejä ovat akateeminen (tutkijat ja opiskelijat), kaupallinen (yritykset), ammatillinen (lääkärit ja opettajat), hallinnollinen (ministeriöt) ja kansalaisyhteiskunta (suuri yleisö).⁷ Toisinaan yhdellä julkaisulla voidaan tavoittaa useita yleisöjä, mutta tieteellisestä julkaisemisesta (*scholarly communication/publishing*) puhuttaessa tarkoitetaan yleensä viestintää tutkijoiden välillä. Tiedeviestinnällä (*science communication/journalism*) tarkoitetaan taas tutkimustulosten julkaisemista ja popularisoimista tiedeyhteisön ulkopuolelle. Käsitteet eivät kuitenkaan aina ole vakiintuneita, ja etenkin keskusteluissa eri käsitteitä käytetään tarkoittamaan eri asioita.

Suomessa opetus- ja kulttuuriministeriö⁸ kerää systemaattisesti tietoa suomalaisten yliopistojen tuottamista julkaisuista (taulukko 1). Tieteelliset julkaisut muodostavat vain yhden, joskin kenties suurimman, kokonaisuuden tiedonkeruussa. Toisinaan voi olla vaikea erottaa tieteellisiä julkaisuja esimerkiksi ammatillisista tai yleistajuisista julkaisuista. Myös opinnäytteisiin, erityisesti väitöskirjoihin, sisältyy tieteellisiä julkaisuja.

Ennen toista maailmansotaa tieteellisen kirjallisuuden julkaiseminen oli keskittynyt lähinnä tieteellisten seurojen toimintaan, jotka julkaisivat oman alansa tieteellistä lehteä, kirjasarjaa tai kirjoja. Sittenmin kustannustoiminta on siirtynyt pääasiassa kaupallisten kansainvälisten kustantajien varaan, joista suurimpia ovat Springer/

6 Garvey 1979; Hagstrom 1965; Meadows 1974.

7 Ylijoki ym. 2011.

8 OKM:n korkeakoulujen tiedonkeruukäsikirja 2022.

4.1 Tieteellinen julkaiseminen – mitä, miksi ja miten?

Taulukko 1. Julkaisutyyppit ja niiden määritelmät opetus- ja kulttuuriministeriön korkeakoulujen tiedonkeruussa vuonna 2022.

Julkaisutyyppi	Määritelmä
Tieteelliset julkaisut	Julkaisut, jotka on tarkoitettu edistämään tiedettä sekä tuottamaan uutta tietoa
Ammatilliset julkaisut	Julkaisut, jotka on tarkoitettu levittämään tutkimukseen perustuvaa tietoa ammattiyhteisön käyttöön
Yleistajuiset julkaisut	Julkaisut, jotka levittävät tieteellistä ja ammatillista tietoa ympäröivään yhteiskuntaan
Taiteelliset julkaisut	Taiteellisen toiminnan tuloksena syntyvät esitykset tai vastaavat tapahtumaluontoiset julkiset tulokset
Opinnäytteet	Ammattikorkeakoulututkinon opinnäytetyö, kandidaatintyö, pro gradu, diplomityö, ylempi amk-opinnäytetyö, lisensiaatintyö, väitöskirja
Muut	Muut korkeakoulujen tutkimustoiminnasta kertovat julkaisut, kuten patentit ja keksintöilmoitukset, audiovisuaaliset julkaisut ja tieto- ja viestintätekniset sovellukset

Nature, Elsevier, Taylor & Francis, Wiley ja Sage. Arvioiden mukaan jopa 50 prosenttia tieteellisistä artikkeleista kansainvälisesti on näiden viiden kustantajan julkaisemia.⁹ Tieteellisellä julkaisutoiminnalla on arvioitu tehtävän liikevaihtoa jopa 30 miljardia dollaria vuosittain, ja sen parissa työskentelee noin 110 000 ihmistä maailmanlaajuisesti.¹⁰ Toisaalta tutkimusta julkaistaan myös kansallisesti. Suomessa voittoa tavoittelemattomat tieteelliset seurat julkaisevat edelleen valtaosan tieteellisistä lehdistä,¹¹ kun taas kaupalliset kustantajat ovat keskittyneet lähinnä tieteellisten kirjojen julkaisemiseen. Suomen kaltaisessa pienen kielialueen maassa tieteellisellä julkaisutoiminnalla voiton tekeminen on haastavaa, mikä selittää myös kaupallisten kustantajien vähäistä roolia Suomessa.

9 Larivière ym. 2015.

10 Johnson ym. 2018.

11 Late ym. 2020.

Tieteelliset julkaisut

Tieteelliset julkaisut ilmestyvät alkuperäisten tutkimustulosten julkaisemiseen erikoistuneissa kanavissa, kuten tieteellisissä lehdissä (*journal*), konferenssijulkaisuissa (*conference proceedings*), kokoomateoksina eli artikkelikokoelmina (*edited work*) ja kirjoina (*monograph*). Alkuperäisyydellä tarkoitetaan sitä, että tulokset julkaistaan ensimmäistä kertaa eli niitä ei ole aiemmin julkaistu muualla. Tieteellisissä julkaisuissa tavoitteena on mahdollisimman yksityiskohtainen kuvaus tutkimuksen toteuttamisesta ja sen tuloksista. Esimerkiksi luonnontieteissä tutkimusmenetelmät ja aineistot pyritään kuvamaan niin tarkasti, että lukija voisi mahdollisuuksien rajoissa toistaa tutkimuksen.

Tieteelliset tekstit näyttävät hieman erilaisilta kuin esimerkiksi sanomalehtiartikkelit. Ensinnäkin tekstit vilisevät viitteitä aiempiin julkaisuihin. Viittausten avulla kirjoittaja sitoo tutkimuksen aiempaan tietoon ja osoittaa, mitä aiheesta toisaalta jo tiedetään ja mitä puutteita tiedossa on. Tieteellisissä teksteissä väitteet pitää perustella ja viittaukset toimivat tässä välineenä. Tieteellisissä teksteissä on tärkeää erottaa omat ajatukset muiden ajatuksista. Toisten ajatusten tai tutkimustulosten esittäminen omanaan on vastoin hyvää tieteellistä käytäntöä ja se luetaan plagioinniksi.¹² Julkaisuihin kohdistuvat viittaukset kertovat julkaisujen vastaanotosta, ja paljon viittauksia saaneita julkaisuja pidetään vaikuttavina. Tieteelliset tekstit myös tyypillisesti noudattelevat tiettyä rakennetta, joka koostuu johdannosta, aiemman tutkimuksen esittelystä, tutkimuskysymyksistä, tutkimusmenetelmistä, tutkimustuloksista ja keskustelusta ja johtopäätöksistä. Tekstin lopusta löytyy aina luettelo lähteistä, joihin tekstissä on viitattu. Toistuva rakenne helpottaa tekstien lukemista ja halutun tiedon paikallistamista.

Tyypillisimmin tutkimustulokset julkaistaan artikkeleina tieteellisissä lehdissä.¹³ Ensimmäiset tieteelliset lehdet perustettiin

12 Tutkimuseettinen neuvottelukunta 2012.

13 Esim. Piro ym. 2013; Puuska & Miettinen 2008.

jo 1600-luvulla, kun Lontoon tieteellisen seuran, Royal Societyn, Philosophical Transactions sekä ranskalainen Journal des Sçavans alkoivat ilmestyä. Tätä ennen tutkimustuloksia jaettiin lähinnä tutkijoiden välisen kirjeenvaihdon kautta.¹⁴ Tänä päivänä englanninkielisiä tieteellisiä lehtiä on arvioitu olevan yli 33 000, ja niiden lisäksi toimii tuhansia muilla kielillä julkaisevia lehtiä.¹⁵ Lehtien tarkkaa määrää on vaikea arvioida, sillä uusia lehtiä syntyy ja vanhoja kuolee rivakkaan tahtiin. Kokonaisuudessaan lehtiä arvioidaan olevan yli 300 000.¹⁶ Tieteellisissä lehdissä julkaistaan vuosittain jopa kolme miljoonaa artikkelia. Suomessa työskentelevät tutkijat julkaisivat vuonna 2020 noin 21 000 tieteellistä lehtiartikkelia, joista yli 80 prosenttia oli englanninkielisiä.¹⁷ Kansainvälisesti Suomen osuus julkaisuista on vain pisara meressä, ja merkittävä osa tieteellisistä artikkeleista tehdään Yhdysvalloissa, Kiinassa ja Intiassa.¹⁸

Konferenssit ovat tutkijoiden tapaamisia, joissa esitellään viimeisimpiä tutkimustuloksia. Osasta konferensseista julkaistaan konferenssijulkaisu, joka koostuu konferenssissa pidettyihin esityksiin pohjautuvista ja julkaistavaksi hyväksytyistä artikkeleista. Kaikki konferenssit eivät ole kuitenkaan tieteellistä julkaisemista varten, vaan pääpaino saattaa useinkin olla keskusteluissa ja verkostojen luomisessa. Lehtien ja konferenssien lisäksi tutkimustuloksia julkaistaan myös kokonaisina kirjoina ja kokoomateoksina (tai antologioina), jotka koostuvat useiden kirjoittajien artikkeleista. Käsillä oleva kirja edustaa tällaista julkaisumuotoa. Väitöskirjat ovat opinnäytteitä, joissa raportoidaan myös alkuperäisiä tutkimustuloksia. Perinteisesti väitöskirja on ollut yksi kokonaisteos, monografia, mutta nykyään mitä useammin väitöskirja koostuu tieteellisissä lehdissä julkaistuista artikkeleista ja näille kirjoitetusta johdannosta.

14 Meadows 1974.

15 Johnson ym. 2018.

16 Forsman 2022.

17 Tiedejatutkimus.fi 29.9.2022.

18 Johnson ym. 2018.

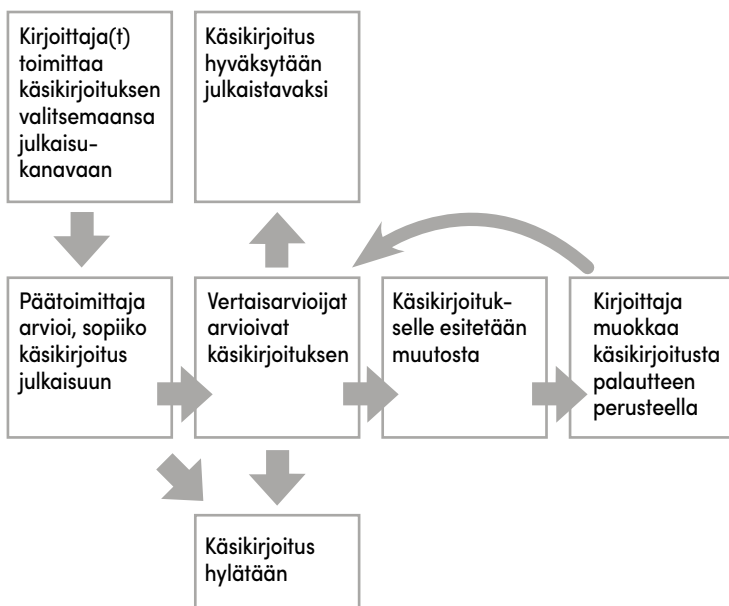
Vertaisarviointi osana julkaisuprosessia

Tieteelliset julkaisukanavat noudattavat tieteellistä julkaisuprosessia (kuva 1), jonka yhtenä erityisominaisuutena on vertaisarviointi (*peer review*). Vertaisarvioinnilla tarkoitetaan menettelyä, jossa tutkimustuloksia julkaiseva lehti, konferenssi tai kirjakustantaja pyytää tieteenalan riippumattomia asiantuntijoita (*reviewer, referee*) arvioimaan julkaistavaksi tarjottua käsikirjoitusta. Prosessin tarkoituksena on varmistaa tulosten paikkansapitävyys, arvioida käsikirjoituksen sopivuutta julkaisukanavaan sekä parantaa käsikirjoituksen laatua. Arvioijia pyydetään pohtimaan esimerkiksi tutkimuksessa käytetyn aineiston kattavuutta, teoreettisen viitekehyksen hallintaa, tutkimuksen toteutuksen luotettavuutta ja tarkkuutta sekä tulosten omaperäisyyttä ja uutuusarvoa suhteessa aiempaan tutkimukseen tieteenalalle ominaisella tavalla. Arvioijat kirjoittavat julkaisun päätoimittajalle käsikirjoituksesta lausunnon, jossa he voivat esittää käsikirjoituksen hylkäämistä (*reject*) tai hyväksymistä sellaisenaan tai vaadituin muutoksin, jotka voivat olla vähäisiä tai huomattavia (*minor/major revisions*). Merkittävä osa käsikirjoituksista hylätään tai niihin vähintäänkin vaaditaan muutoksia ennen julkaisua.¹⁹

Arvioijien riippumattomuuden varmistamiseksi arvioijat toimivat usein anonyymeinä. Kun puhutaan *double blind* -arvioinnista (kaksoissokkoarviointi), tarkoitetaan tilannetta, jossa sekä kirjoittajan että arvioijan henkilöllisyys salataan vertaisarviointiprosessissa. *Single blind* -arvioinnissa (yksöissokko) vain arvioija toimii anonyymisti. Anonyymisyys ei kuitenkaan aina takaa arvioijien puolueettomuutta, ja etenkin pienissä maissa ja pienillä tieteenaloilla tutkijat saattavat tunnistaa toistensa työn anonymisoinnista huolimatta. Toisinaan vertaisarvioinnissa on huomattu syrjintää esimerkiksi kirjoittajan iän, sukupuolen, uravaiheen tai edustaman organisaation tai maan perusteella. Yhtenä vertaisarvioinnin ongelmana on

19 Weller 2001.

4.1 Tieteellinen julkaiseminen – mitä, miksi ja miten?



Kuva 1. Vertaisarviointi osana tieteellistä julkaisuprosessia. Lähde: Kansallinen julkaisumetriikkaopas, 2022 (<https://wiki.eduuni.fi/display/cscckjmo/Vertaisarviointi>). CC BY-SA 4.0.

myös sen hitaus, sillä toisinaan tieteelliset julkaisuprosessit kestävät jopa vuosia.²⁰

Tieteenalojen erilaiset julkaisukäytännöt

Tieteenaloilla on omat tapansa työskennellä ja tuottaa tietoa. Puhutaan tieteenalakulttuureista, jotka määrittävät muun muassa alan tapaa tuottaa tietoa ja viestiä siitä.²¹ Tieteenalaksi nimitetään tietämyksen aluetta, jolla on oma, sille ominainen tutkimusaihe ja oma tai omia tutkimusmenetelmiä ja jonka piiriin kuuluvat tutkijat tiedos-

²⁰ Weller 2001.

²¹ Esim. Becher & Trowler 2001.

tavat olevansa kyseisen tieteenalan tutkijoita. Tieteenalat ovat usein myös järjestäytyneet esimerkiksi yliopistojen oppiaineiksi. Tieteenalalla on lisäksi oma tieteellinen yhteisönsä, joka mahdollisesti kuuluu sen nimeä kantavaan seuraan, julkaisee omaa aikakauslehteä ja järjestää sen nimissä esitelmä- ja koulutustilaisuuksia, seminaareja ja konferensseja. Tieteenaloja ovat esimerkiksi matematiikka, fysiikka, terveystiede, rakennustekniikka, tietotekniikka, viestintätiede ja historia. Laajemmiksi tieteenalaryhmiksi on määritelty luonnontieteet, tekniikka, lääke- ja terveystieteet, maatalous- ja metsätieteet, yhteiskuntatieteet ja humanistiset tieteet.²² Tutkimukselle ei aina pystytä määrittämään vain yhtä tieteenalaa, vaan yhä useammin tutkimuksessa yhdistyy usean eri tieteenalan piirteitä. Puhutaan monitieteisyydestä, tieteidenvälisyydestä ja poikkitieteellisyydestä.

Luonnontiedettä, lääketiedettä ja teknillisiä aloja on kuvailtu perinteisesti koviksi aloiksi, joissa tutkimus on kumulatiivista ja eksaktia ja pyrkii yleisten lainalaisuuksien selvittämiseen. Yhteiskuntatieteitä ja humanistisia aloja kuvataan usein pehmeiksi aloiksi, joilla tutkimus keskittyy yksittäisten ilmiöiden tarkasteluun ja käytetyt teorit ja menetelmät ovat moninaisia. Aloja erottaa myös tutkimuksen soveltavuus eli se, missä määrin tutkimusta voidaan suoraan hyödyntää tieteen ulkopuolella.²³ Tieteenalojen julkaisukäytäntöihin vaikuttavat vahvasti alan viestintäkieli, tutkijoiden riippuvuus toisistaan, tutkimuksen yleisöt, tutkimuskohteiden paikallisuus tai globaalius ja tutkimusalalla vallitseva kilpailu.²⁴

Tieteenalakulttuurit vaikuttavat siihen, missä ja miten tutkimustuloksia julkaistaan. Vaikka yleisesti ottaen englanninkieliset artikkelit tieteellisissä lehdissä ovat tyypillisin julkaisumuoto kaikilla aloilla, alojen julkaisukäytännöissä on merkittäviä eroja.²⁵ Luonnon- ja lääketieteiden aloilla kansainväliset tieteelliset lehtiartikkelit ovat vakiintunut julkaisumuoto, kun taas esimerkiksi teknillisillä

22 OECD 2015; Finto 2022.

23 Becher & Trowler 2001.

24 Kyvik 1999; Puuska & Miettinen 2008.

25 Esim. Puuska & Miettinen 2008; Sivertsen 2019.

aloilla konferenssijulkaiseminen on yleistä.²⁶ Yhteiskuntatieteissä ja humanistisilla aloilla julkaisukäytännöt ovat moninaisemmat. Näillä aloilla julkaiseminen kansallisilla foorumeilla, kansallisilla kielillä sekä kirjajulkaiseminen on selvästi yleisempää verrattuna muihin aloihin.²⁷

Tieteenaloille ominaiset työskentelytavat, kuten työskennelläänkö suurissa tutkimusryhmissä vai yksin, vaikuttavat myös julkaisukäytäntöihin. Luonnon- ja lääketieteissä tutkimuksen tekeminen vaatii usein suuren tutkijajoukon. Julkaisuissa saattaakin olla satoja ja toisinaan jopa tuhansia kirjoittajia. Näillä aloilla myös julkaisutempo on tiheämpi, ja näin ollen myös viittauksia julkaisuihin kertyy enemmän. Humanistisilla ja yhteiskuntatieteellisillä aloilla työskentely tapahtuu useimmiten pienissä ryhmissä ja yksinkirjottaminen on yleistä. Yhteistyön määrä kuitenkin lisääntyy kaikilla aloilla.²⁸ Olennaista tieteellisen julkaisemisen tarkastelussa onkin ymmärtää eri tieteenalojen tapoja tuottaa ja julkaista tietoa.

Julkaisukanavien arviointi

Vaikka tieteellisillä julkaisukanavilla on yhteneväiset julkaisuprosessit, niiden arvostus julkaisufoorumeina vaihtelee. Kullakin tieteenalalla on omat arvostetuimmat julkaisukanavansa, joissa julkaistaan ainoastaan alan huippututkimusta. Esimerkiksi myös suuren yleisön hyvin tuntemissa Nature- ja Science-lehdissä julkaisemista arvostetaan suuresti, ja ne valikoivat tarkasti, mitä lehtiin hyväksytään julkaistavaksi. Sama pätee myös konferensseihin ja kirjajulkaisijoihin. Arvostetuimmissa kanavissa julkaisemista pidetään tutkijan meriitinä, ja nämä julkaisut saavat usein myös eniten huomiota ja myös viittauksia muissa julkaisuissa.

Julkaisukanavien laadunarviointia varten Suomeen perustettiin vuonna 2010 Julkaisufoorumi (Jufo), joka ylläpitää julkaisu-

26 Piro ym. 2013; Puuska & Miettinen 2008.

27 Kulczycki ym. 2020; Pölönen 2021; Sivertsen 2019.

28 Puuska & Miettinen 2008.

kanavien tasoluokitusta. Julkaisufoorumissa toimii 23 tieteenala-kohtaista asiantuntijapaneelia, jotka luokittavat alansa tieteellisiä lehtiä, konferensseja ja kirjajulkaisijoita kolmelle eri tasolle. Taso 1 on perustaso, kun taas tasolle 3 luokitetaan alan johtavat kansainväliset julkaisukanavat. Tasolle 1 hyväksytään käytännössä kaikki Jufon kriteerit täyttävät vertaisarvioidut tieteelliset kanavat. Julkaisufoorumiluokitus on sidottu opetus- ja kulttuuriministeriön yliopistoille myöntämään perusrahoitukseen siten, että 14 prosenttia rahoituksesta jaetaan julkaisujen ja niiden laadun perusteella. Toisin sanoen mitä enemmän tutkijat julkaisevat ylempien tasojen julkaisukanavissa, sitä enemmän yliopisto saa rahoitusta. Jufo on saanut kritiikkiä erityisesti humanistisilla ja yhteiskuntatieteellisillä aloilla, sillä sen on koettu arvottavan korkeammalle kansainvälisiä kanavia, vaikka joillakin aloilla kotimaassa julkaiseminen on luontevampaa.²⁹ Tanskassa ja Norjassa on päätetty luopua vastaavan tasoluokituksen käytöstä, ja myös Suomessa Jufon tulevaisuutta tarkastellaan.³⁰

Julkaisukanavien arviointiin on kehitetty myös erilaisia määrällisiä mittareita. Näistä tunnetuin on tieteellisten lehtien vaikuttavuuskerroin (*Journal Impact Factor*, JIF), joka osoittaa kuinka monta viittautusta lehdessä julkaistu artikkeli keskimäärin saa. Korkean luvun katsotaan kertovan lehden laadusta ja vaikuttavuudesta. JIF vaihtelee merkittävästi tieteenalojen välillä. Esimerkiksi Nature-lehden kerroin on lähes 70, kun taas yhteiskuntatieteissä korkeimmat kertoimet ovat huomattavasti matalampia. Myös JIF on saanut runsaasti kritiikkiä, sillä sen laskeminen perustuu vain yhteen viittautustietokantaan (Web of Science) indeksoituihin lehtiin. Suurelle osalle lehtiä ei siis voida laskea vaikuttavuuskerrointa. Lehdet voivat myös pyrkiä kasvattamaan kerrointa esimerkiksi valikoimalla julkaistavaksi paljon viittauksia kerääviä artikkeleita ja ohjaamalla kirjoittajia viittamaan samassa lehdessä julkaistuun sisältöön. Vaikuttavuuskerroin ei myöskään voi ottaa huomioon viittausten laatua

29 Esim. Rantala 2012.

30 Pölönen 2022; Pölönen 2023.

eli sitä, tukevatko vai kritisoivatko ne tutkimuksia. Ylipäättään JIF ja vastaavan kaltaiset mittarit rajaavat julkaisujen vaikuttavuuden vain viittauksiin, vaikka toisinaan viittaukset julkaisuihin voivat kertyä hyvinkin pitkän ajan kuluessa ja toisaalta vaikuttavuus voi ilmetä myös muulla tavalla. Puutteistaan huolimatta vaikuttavuuskerroin on yleisesti käytössä erityisesti luonnon- ja lääketieteen aloilla. JIF:n lisäksi on kehitetty muitakin julkaisujen vaikuttavuuden indikaattoreita, jotka pyrkivät huomioimaan paremmin tieteenalaeroja.³¹

Tieteellinen julkaiseminen muutoksessa

Digitalisaatio muutti merkittävästi tieteellisen tiedon julkaisemista 1990-luvulla, kun erityisesti lehdet siirtyivät ensin digitaaliseen ja pian internetistä saatavaan muotoon. Kun vielä 20 vuotta sitten tutkijat käyttivät paljon aikaa kirjastoissa etsien ja lukien tieteellisiä lehtiä ja kirjoja, nykypäivän tutkija saa suuren osan tarvitsemastaan tiedosta internetin kautta.³² Tieteelliset kirjastot ovat mahdollistaneet tutkijoille ja opiskelijoille pääsyn tieteelliseen tietoon hankkimalla lisenssejä lehtien lukemiseen. Suomessa on 1990-luvulta saakka toiminut FinELib-konsortio, joka on valtakunnallisten konsortiosopimusten kautta neuvotellut pääsyn erilaisiin digitaalisiin aineistoihin, kuten tietokantoihin, lehtiin, kirjoihin ja kokoelmiin. Samaan aikaan digitalisaation kanssa tieteellisten lehtien tilausmaksut lähtivät nousuun synnyttäen niin sanotun lehtikriisin (*serials crises*).³³ Kaupallisten kustantajien tehdessä voittoja tieteellisen kirjallisuuden julkaisemisella yliopistokirjastojen budjetit eivät enää riittäneet kasvaviin lisenssi- ja tilausmaksuihin. Tilannetta on pidetty kohtuuttomana, sillä julkaisujen sisältö on pääasiassa valtion rahoittamien tutkimusorganisaatioiden tuottamaa. Myös julkaisuprosessiin osallistuvat vertaisarvioijat tekevät työn osana muuta työtään, eikä siitä yleensä makseta erillisiä palkkioita.

31 Ks. Forsman 2022.

32 Late ym. 2019.

33 Panitch & Michalak 2005.

Ainakin osittain reaktiona tähän syntyi ajatus avoimesta julkaisemisesta, jonka lähtökohtana on sisältöjen maksuttomuus lukijoille. Eri organisaatiot, tutkimusrahoittajat, valtiot ja Euroopan unioni ovat painottaneet tiedon avoimuutta 2000-luvun aikana.³⁴ Esimerkiksi Suomeen on perustettu Avoimen tieteen koordinaatio,³⁵ joka edistää ja ohjaa Suomessa annettuja avoimen tieteen linjauksia.³⁶ Myös eurooppalaisten tutkimusrahoittajien (mukaan lukien Suomen Akatemia) muodostaman konsortion vuonna 2018 allekirjoittama sopimus Plan S edellyttää rahoituksen saaneita tutkijoita avoimeen julkaisemiseen.³⁷

Avoim julkaiseminen

Tieteen avoimuus ei sinänsä ole mikään uusi keksintö, vaan osana esimerkiksi jo 1960-luvulla Robert Mertonin esittämiä tieteen normeja. Digitalisoituminen on kuitenkin avannut uusia mahdollisuuksia avoimuudelle. Avoimella julkaisemisella (*open access*, OA) tänä päivänä tarkoitetaan julkaisujen vapaata saatavuutta ilman tilausmaksuja. Avoimuus voidaan toteuttaa eri tavoin. Julkaisu voi olla lähtökohtaisesti avoimesti saatavilla ja kenen tahansa luettavana tai julkaisun versio voidaan tallentaa avoimeen julkaisuarkistoon, josta se on vapaasti saatavilla. Tätä kutsutaan rinnakkaistallentamiseksi. Toisinaan julkaisut ovat avoimesti luettavissa vasta viiveen jälkeen. Useimmat yliopistot ja muut tutkimusorganisaatiot ovat perustaneet omat julkaisuarkistonsa rinnakkaistallennusta varten. Olemassa on myös tieteenalakohtaisia ja tieteenalarajat ylittäviä arkistoja, joista esimerkkejä ovat Arxiv ja Zenodo. Näihin arkistoihin voidaan tallentaa muualla julkaistuja tai vielä julkaisemattomia materiaaleja tekijänoikeuksien sallimissa rajoissa.

34 Unesco 2021.

35 avointiede.fi

36 Avoimen tieteen koordinaatio, Tieteellisten seurain valtuuskunta 2020.

37 Coalition S 2025.

Yhä useampi tieteellinen lehti toimii nykyään avoimesti.³⁸ Esimerkiksi Suomessa toimivien tutkijoiden vuosina 2018–2020 julkaisemista vertaisarvioituista julkaisuista yli 50 prosenttia oli vapaasti luettavissa. Avoimuus on yleisempää artikkelien osalta, kun taas kirjoista pienempi osa on avoimia.³⁹ Directory of open access journals (DOAJ) on avointen tieteellisten lehtien rekisteri, johon syksyllä 2024 on rekisteröity 21 205 tieteellistä lehteä. Luku ei kuitenkaan kerro koko totuutta avointen lehtien määrästä, sillä kaikki lehdet eivät ole rekisteröityneet tietokantaan. Muutos on synnyttänyt myös uudenlaisia julkaisukanavia ja niin kutsuttuja megalehtiä (*mega journal*). Megalehdillä tarkoitetaan suuria määriä artikkeleita julkaisevia lehtiä tai julkaisualustoja, jotka eivät perinteisten tieteellisten lehtien tapaan ole fokusoituneita tietylle tieteentalalle. Kenties tunnetuin näistä on PLOS ONE.⁴⁰

Ajatus tieteen avoimuudesta on herättänyt ajatuksia myös vertaisarvioinnin kehittämiseen. Avoimuus vertaisarvioinnissa voi koskea kirjoittajien ja arvioijien identiteettejä ja/tai arviointiläisyyksiä.⁴¹ Osa julkaisuista pyrkii osallistamaan koko lukijakunnan arviointiin ja suorittamaan arvioinnin julkaisemisen jälkeen (esimerkiksi Open Research Europe ja F1000). Jo vuosia erityisesti luonnontieteiden aloilla artikkeleiden *preprint*-versioiden (ennen vertaisarviointia tai julkaisupäätöstä) julkaiseminen on yleistynyt. Koska keskeistä vertaisarvioinnille perinteisesti on ollut arvioijien riippumattomuus ja anonyymiyys, kaikki tutkijat eivät ole halukkaita toimimaan arvioijina silloin, kun arviointi ei ole nimetöntä. Monet julkaisijat ovat kokeneet vaikeuksia vertaisarvioijien rekrytoimiseen perinteisessä mallissa, ja pelkona on, että ongelma kasvaa avoimessa arvioinnissa. Avoimuus vertaisarvioinnissa on toistaiseksi kuitenkin vielä harvinaista.⁴²

38 Björk 2019.

39 Pölönen 2021.

40 Spezi ym. 2017.

41 Ross-Hellauer 2017.

42 Jytilä & Laakso 2019.

Tieteen avoimuuteen liittyy myös kysymys tutkimusaineistojen avoimuudesta. Yhä useampi tieteellinen lehti edellyttää tai ainakin suosittelee tutkimusten takana olevien tutkimusaineistojen avointa jakamista.⁴³ Aineistoja voidaan tallentaa niiden jakamiseen erikoistuneisiin data-arkistoihin. Näistä yhtenä esimerkkinä on Suomessa toimiva Tietoarkisto, joka tallentaa ja tarjoaa pääsyt erityisesti yhteiskuntatieteellisiin aineistoihin. Viime vuosina uusiksi julkaisumuodoiksi ovat vakiintuneet niin sanotut data-artikkelit, jotka tutkimustulosten sijaan julkaisevat tietoa tutkimusaineistosta. Data-artikkeleiden julkaisemiseen on syntynyt niiden julkaisemiseen erikoistuneita lehtiä.⁴⁴

Avoimuuden hinta

Vaikka ainakin osittain pyrkimys avoimeen julkaisemiseen liittyi tieteellisten lehtien kasvaviin tilaus- ja lisenssimaksuihin, avoin julkaiseminen ei ole toistaiseksi tuottanut säästöjä tutkimusorganisaatioille,⁴⁵ vaan siirtyminen avoimeen julkaisemiseen on muuttanut julkaiseminen ansaintalogiikkaa. Kun aiemmin julkaisujen lukijat ja/tai kirjastot maksoivat sisällöistä, avoimen julkaisemisen myötä lasku on siirtynyt kirjoittajalle. Monet avoimet lehdet perivät kirjoittajamaksun (*author processing charges*, APC), jolla katetaan julkaisemisen kustannukset. Kirjoittajamaksu voi olla useita tuhansia euroja riippuen julkaisusta. Osa lehdistä tarjoaa kirjoittajille mahdollisuutta ostaa artikkeli avoimeksi perinteisesti toimivasta lehdestä. Näitä lehtiä kutsutaan hybridilehdiksi. Kirjastojen kustantajien kanssa solmimat sopimukset koskevatkin yhä useammin kirjoittajamaksuja, jolloin kirjoittajamaksut katetaan organisaation sopimuksilla. Voittoa tavoittelevan tieteellisen julkaisutoiminnan ohella toimii useita ei-voittoa tavoittelevia avoimia julkaisukanavia, jotka eivät yleensä peri kirjoittajamaksuja. Tällaisia tieteellisiä lehtiä

43 Jones ym. 2019; Rousi & Laakso 2020.

44 Candela ym. 2015.

45 Laakso 2021; Stoy ym. 2019.

kutsutaan timanttilehdiksi (*Diamond open access*), ja ne rahoittavat toimintaansa erilaisin avustuksin ja toimivat vapaaehtoistyövoimin.⁴⁶ Esimerkiksi merkittävä osa suomalaisista avoimista tieteellisistä lehdistä toimii tähän tapaan. Suomalaisia tieteellisiä lehtiä voi lukea journal.fi-sivuston kautta.

Avoimeen julkaisemiseen siirtyminen ei ole ollut täysin ongelmattonta. Julkaisemisen ansaintalogiikan muuttumisen myötä on syntynyt niin kutsuttuja saalistajalehtiä (*predator journals*), jotka eivät noudata tieteellisen julkaisemisen eettisiä sääntöjä esimerkiksi vertaisarvioinnin osalta, vaan pyrkivät julkaisemaan mahdollisimman paljon artikkeleita kerätäkseen voittoja kirjoittajamaksuilla.⁴⁷ Näitä lehtiä on pyritty kartoittamaan, ja niiden tunnistamiseksi on luotu erilaisia listauksia.⁴⁸ Huomiota ovat saaneet myös raportit testeistä, joissa tekaistuja artikkeleita on lähetetty julkaistavaksi epäilyksen alaisiin lehtiin ja jopa sadat lehdet ovat antaneet hyväksyvän julkaisupäätöksen.⁴⁹ Epäeettisesti toimivat lehdet ovatkin merkittävä uhka tieteen uskottavuudelle. Toisinaan niin sanottuja saalistajalehtiä on vaikea tunnistaa, sillä osa kustantajista saattaa kustantaa sekä eettisesti että vastuuttomasti toimivia julkaisuja.⁵⁰ Julkaisevien tutkijoiden onkin oltava tarkkana valitessaan luotettavia ja laadukkaita julkaisukanavia.

Informetriikka tieteellisen julkaisemisen tutkimusmenetelmänä

Tieteellisen julkaisutoiminnan tutkimus on oma vakiintunut tutkimusala, osana informaatiotutkimusta ja tieteen ja teknologian tutkimusta. Julkaisutoiminnan tutkimus ei kuitenkaan ole rajoittunut ainoastaan näille aloille, vaan eri tieteenaloilla, kuten lääketie-

46 Bosman ym. 2021.

47 Grudniewicz ym. 2019.

48 Koerber ym. 2020.

49 Bohannon 2013.

50 Crosetto 2021.

teissä, tarkastellaan oman alan julkaisukäytäntöjä, tuotteliaimpia tai viitatuimpia tutkijoita tai vaikkapa vaikutusvaltaisimpia julkaisukanavia. Tieteenalakohtaiset erot tiedon tuotannon ja julkaisemisen tavoissa onkin aina syytä ottaa huomioon tämän alueen tutkimuksessa. Tieteellisen julkaisemisen tutkimuksella pyritään ymmärtämään tieteellisen tiedon tuotannon ja käytön ilmiöitä.

Informetrinen tutkimus on yksi perinteisimmistä tutkimusmenetelmien kokonaisuuksista, joilla julkaisu- ja viittaustoimintaa on tarkasteltu. Informetrisillä menetelmillä tarkoitetaan määrällistä tutkimusta, jossa tutkimusaineistona hyödynnetään tietokannoista koottuja julkaisujen bibliografisia tietoja ja viittautustietoja. Informetrisiin menetelmiin kuuluvat bibliometriikka (julkaisuihin kohdistuva tutkimus), scientometriikka (tieteellisiin julkaisuihin kohdistuva tutkimus) ja webometriikka (verkkojulkaisuihin kohdistuva tutkimus). Suomen kielessä toisinaan käytetään myös nimitystä julkaisumetriikka.⁵¹ Bibliometriikan käsite on näistä vakiintunein, ja sitä käytetään yleisesti viittamaan eri informetrisiin perinteisiin.

Informetrisin menetelmin tarkastellaan tyypillisesti julkaisujen määrää, tyyppiä, kirjoittajia, viittauksia tai käytettyjä lähteitä. Viittausanalyysi on yksi informetriikan suosituimmista menetelmistä. Siinä tarkastellaan julkaisun saamien viittauksien määrää ja alkuperää. Lähdeanalyysissa tarkastelun kohteena on julkaisun käyttämät lähteet. Julkaisuihin voidaan analysoida myös käytettyjä käsitteitä esimerkiksi yhteissana-analyysilla. Informetriselle tutkimukselle on tyypillistä myös sellaisten indikaattoreiden käyttö ja luominen, joiden avulla pyritään kuvaamaan vaikkapa tuottavuutta, vaikuttavuutta tai yhteistyötä. Esimerkiksi Forsman⁵² ja Kortelainen⁵³ ovat esitelleet erilaisia analyysseja ja niiden peruseriaatteita yksityiskohteisesti. Vuonna 2022 julkaistu vapaasti saatavilla oleva Kansallinen julkaisumetriikkaopas on myös hyödyllinen väline alan tutkimusta tekeväälle.

51 Kansallinen julkaisumetriikkaopas 2022.

52 Forsman 2022.

53 Kortelainen 2010.

Tutkimusaineistot informetriikassa

Julkaisu- ja viittaussaineistot tutkimusta varten kerätään yleensä julkaisutietokannoista, joista löytyy julkaisujen bibliografisten tietojen (esimerkiksi julkaisun nimi, kirjoittajat, julkaisukanava ja julkaisuaika) lisäksi tiedot julkaisuun viittaavista muista julkaisuista. Tietokannoista käytetyimpiä ovat Web of Science (Wos, Thomson Reuters) ja Scopus (Elsevier), joita on ylläpidetty useita vuosia ja joiden sisältö on kuratoitua ja päivittyvää. Tietokantoihin indeksoidaan eri tieteenalojen keskeisimpiä julkaisukanavia rajoittuen pääasiassa englanninkielisiin tieteellisiin lehtiin. Tietokannat tarjoavat myös helppokäyttöisiä analyysityökaluja julkaisujen tarkasteluun ja tulosten visualisointiin. Tietokantojen käyttö edellyttää lisenssiä, joka on saatavilla useiden yliopistokirjastojen kautta.

Näiden lisäksi tutkimusaineistoja voidaan kerätä maksuttomasti esimerkiksi Dimensions (Digital science), Open Alex ja Google Scholar -tietokannoista. Verrattuna muihin tiedonlähteisiin Google Scholarin kattavuus on hyvä,⁵⁴ sillä se indeksoi eri kielisiä ja tyyppisiä julkaisuja. Google Scholarin käyttö tutkimustarkoituksiin on kuitenkin hankalaa, sillä ei ole täysin selvää, mitä kaikkea Google Scholariin indeksoidaan. Google Scholarista löytyy varsinaisten tieteellisten julkaisujen lisäksi myös muun tyyppisiä julkaisuja ja julkaisuiden eri versioita. Kattavuutensa vuoksi Google Scholar onkin erittäin käytetty tietokanta julkaisujen hakemiseen, mutta informetrisessä tutkimuksessa sen kautta hankittu aineisto vaatii huomattavasti enemmän työtä kuin maksullisista tietokannoista kerätty aineisto. Aineistosta täytyy poistaa päällekkäiset julkaisut sekä ei-tieteelliseksi julkaisuiksi (esimerkiksi ei-vertaisarvioidut) luokitellut julkaisut. Google Scholarin tarjoamat viittaustiedot voivat olla myös epäluotettavia. Koska tutkimusaineistot ovat yleensä suuria ja käsittävät tuhansia julkaisuja, tehtävä on työläs.

Eri maat ja tutkimusorganisaatiot ylläpitävät myös paikallisia tietokantoja, joista löytyy tietyn maan tai organisaation julkaisut.

54 Esim. Martín-Martín ym. 2021.

Näitä kutsutaan usein myös CRIS-tietokannoiksi (*Current research information system*). Suomessa tuotettujen julkaisujen tiedot löytyvät Virta-julkaisupalvelusta. Paikallisista tietokannoista löytyy todennäköisemmin kyseisen organisaation tai maan kaikki julkaisutuotanto, kun taas kaupalliset tietokannat valikoivat sisältönsä julkaisukanavien perusteella. Esimerkiksi WoS ja Scopus indeksoivat tietokantoihinsa pääasiassa englanninkielisiä lehtiartikkeleita. Luonnon- ja lääketieteiden osalta kaupallisten tietokantojen kattavuus julkaisutuotannosta on melko hyvä, kun taas humanististen ja yhteiskuntatieteiden osalta tietokannat kattavat vain osan tutkijoiden julkaisuista.⁵⁵ Toisaalta paikallisista tietokannoista ei löydy julkaisujen saamia viittauksia, joten niiden käyttö on rajatumpaa. Informetristä tutkimusta tehdessä on tärkeää ymmärtää käytetyn tietokannan tiedonlähteet ja niiden rajoitukset. Ylipäättään analyysissä täytyy ymmärtää tieteenalakohtaiset julkaisukulttuurit eikä tieteenaloja voida vertailla keskenään samoja indikaattoreita käyttäen.

Tulevaisuuden suuntia

Tieteellinen julkaiseminen muuttuu digitalisaation ja uusien tiedepoliittisten linjausten myötä. Ajankohtaisia teemoja ovat tieteen avoimuus, vastuullisuus ja yhteiskunnallinen vaikuttavuus. Tieteellisessä julkaisemisessa ja sen tutkimuksessa tämä näkyy pyrkimyksessä laajempaan ymmärrykseen tieteellisestä viestinnästä. On selvää, että keskittyminen kansainvälisiin vertaisarvioituihin julkaisuihin jättää väistämättä huomiotta osan julkaisutoiminnasta. Esimerkiksi kansallisilla kielillä julkaisemisen merkitys on viime vuosina noussut keskusteluun.⁵⁶ Tieteelliset julkaisut eivät myöskään tavoita kaikkia tutkimuksen yleisöjä, joten tutkimuksesta pitää viestiä myös toisin tavoin, tiedeviestinnän keinoin. Tätä on toki aina tehty, mutta

55 Mongeon & Paul-Hus 2016; Sivertsen 2016.

56 Esim. Helsinki Initiative on multilingualism in scholarly communication 2019.

sen merkitys saattaa tulevaisuudessa korostua. Kehitys näkyy myös tieteellisissä julkaisuissa, joihin yhä useammin liitetään sisältöjä popularisoivia osia, kuten videoita, joissa kirjoittajat voivat kertoa tutkimuksestaan.

Keskusteluun on noussut myös kysymys tutkimuksen arvioinnin vastuullisuudesta. Monissa kansallisissa ja kansainvälisissä julkilausumissa ja ohjeissa on nostettu esiin erilaisten tutkimustuotosten ja tutkijoiden tehtävien aiempaa monipuolisempi huomioiminen arvioinnissa sekä osoitettu käytössä olevien metriikoiden ja indikaattoreiden ongelmia.⁵⁷ Tutkimuksen vastuullisessa arvioinnissa korostetaan myös laadullisen asiantuntija-arvioinnin merkitystä, jolloin tilastolliset analyysit ovat vain osa arviointia.

Tutkimuksen yhteiskunnallista vaikuttavuutta julkaisujen näkökulmasta on pyritty selvittämään informetrisin menetelmin. Kiinnostavaa on, miten paljon tieteellisissä julkaisuissa raportoituja tuloksia hyödynnetään tutkimuksen ulkopuolella, esimerkiksi politiikassa, jossa yhtenä pyrkimyksenä on tietoon perustuva päätöksenteko. Viestinnän ja tutkimusaineistojen digitalisoituminen on mahdollistanut uusien tutkimusasetelmien ja menetelmien käytön tieteellisen julkaisemisen tutkimuksessa. Tästä yhtenä esimerkkinä on altmetriikka, jonka avulla pyritään selvittämään julkaisujen käyttöä ja vaikuttavuutta ja huomiota perinteisten tieteellisten kanavien ulkopuolella, kuten sosiaalisessa mediassa.⁵⁸

Vaikka informetriset menetelmät ovat suosittuja tieteellisen julkaisemisen tutkimuksessa, tutkimusta toteutetaan myös muilla menetelmillä. Julkaisuaineistot eivät vastaa kysymyksiin vaikkapa kirjoittajien käytännöistä ja kokemuksista, jotka liittyvät julkaisemiseen tai julkaisujen lukemiseen.⁵⁹ Tähän tarvitaan usein laadullisia lähestymistapoja, joita voidaan myös yhdistää määrällisiin tulok-

57 Esim. European Commission 2021; San Francisco declaration of research assessment 2010; Vastuullisen tutkijanarvioinnin työryhmä 2020; Wilsdon ym. 2015.

58 Esim. Forsman 2022.

59 Esim. Talja ym. 2021.

siin selittämään tuloksia ja tuomaan syvempää ymmärrystä niihin. Esimerkiksi perinteisillä kysely- ja haastattelumenetelmillä voidaan kerätä tietoa eri julkaisuprosessiin osallistuvilta tahoilta, kuten kirjoittajilta, vertaisarvioijilta, toimittajilta, julkaisijoilta ja kirjastoilta. Eri lähestymistapojen ja tutkimusmenetelmien käyttö laajentaa ymmärrystä tieteellisestä julkaisemisesta ja sen kehityksestä.

Lähteet

- Avoimen tieteen koordinaatio, Tieteellisten seurain valtuuskunta (2020) *Avoimen tieteen ja tutkimuksen julistus 2020–2025*. Vastuullisen tieteen julkaisusarja 1:2020. Tiedonjulkistamisen neuvottelukunta ja Tieteellisten seurain valtuuskunta. <https://doi.org/10.23847/isbn.9789525995237>
- Becher, Tony & Paul Trowler (2001) *Academic tribes and territories: intellectual enquiry and the cultures of disciplines*. Lontoo: The Society for Research into Higher Education & Open University Press.
- Björk, Bo-Chirster (2019) Open access journal publishing in the Nordic countries. *Learned publishing*, 32:3, 227–236. <https://doi.org/10.1002/leap.1231>
- Bohannon, John (2013) Who's afraid of peer review? *Science*, 342:6154, 60–65. <https://doi.org/10.1126/science.342.6154.60>
- Borgman, Christine (2007) *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/7434.001.0001>
- Bosman, Jeroen & Jan Erik Frantsevåg & Bianca Kramer & Pierre-Carl Langlais & Vanessa Proudman (2021) *The OA diamond journals study. Part 1: Findings*. <https://doi.org/10.5281/zenodo.4558704>
- Candela, Leonardo & Donatella Castelli & Paolo Manghi & Alice Tani (2015) Data journals: A survey. *Journal of the association for information science and technology*, 66:9, 1747–1762. <https://doi.org/10.1002/asi.23358>
- Coalition S (2025) <https://www.coalition-s.org/> (haettu 18.2.2025).
- Crosetto, Paolo (2021) *Is MDPI a predatory publisher?* <https://paolocrosetto.wordpress.com/2021/04/12/is-mdpi-a-predatory-publisher/>
- Directory of open access journals. <https://doaj.org/>
- European commission, directorate-general for research and innovation (2021) *Towards a reform of the research assessment system: scoping report*, Publications Office. <https://data.europa.eu/doi/10.2777/707440>
- Finto (2022) *Korkeakoulujen tutkimustiedonkeruussa käytettävä tieteenalaluokitus*. <http://www.yso.fi/onto/okm-tieteenala/conceptscheme>
- Forsman, Maria (2022) *Julkaisut ja tieteen mittaaminen bibliometriikan käännekohtia*. Toinen, tarkastettu painos. Helsinki: Enostone kustannus. Teos on julkaistu Helsingin yliopiston Helda Open Books -kokoelmassa.

4.1 Tieteellinen julkaiseminen – mitä, miksi ja miten?

- https://helda.helsinki.fi/bitstream/handle/10138/343967/Forsman_Julkaisut_ja_tieteellinen_mittaaminen_2022.pdf
- Hagstrom, Warren (1965) *The scientific community*. New York: Basic books.
- Helsinki initiative on multilingualism in scholarly communication (2019) Helsinki: Federation of Finnish learned societies, committee for public information, Finnish association for scholarly publishing, universities Norway & European network for research evaluation in the social sciences and the humanities. <https://doi.org/10.6084/m9.figshare.7887059>.
- Garvey, William (1979) *Communication, the essence of science: Facilitating information exchange among librarians, scientists, engineers, and students*. Oxford: Pergamon Press.
- Grudniewicz, Agnes & David Moher & Kelly Cobey & Gregory Bryson & Samantha Cukier & Kristiann Allen & Clare Arden & Lesley Balcom & Tiago Barros & Monica Berger & Jairo Buitrago Ciro & Lucia Cugusi & Michael Donaldson & Matthias Egger & Ian Graham & Matt Hodgkinson & Karim Khan & Mahlubi Mabizela & Andrea Manca & Katrin Milzow & Johann Mouton & Marvelous Muchenje & Tom Olijhoek & Alexander Ommaya & Bhushan Patwardhan & Deborah Poff & Laurie Proulx & Marc Rodger & Anna Severin & Michaela Strinzel & Mauro Sylos-Labini & Robyn Tamblyn & Marthie van Niekerk & Jelte Wicherts & Manoj Lalu (2019) Predatory journals: no definition, no defence. *Nature* 576, 210–212. <https://doi.org/10.1038/d41586-019-03759-y>
- Johnson, Rob & Anthony Watkinson & Michael Mabe (2018) *The STM Report: An overview of scientific and scholarly publishing*. https://policycommons.net/artifacts/1575771/2018_10_04_stm_report_2018/2265545/
- Jones, Leila & Rebecca Grant & Iain Hrynaszkiewicz (2019) Implementing publisher policies that inform support and encourage authors to share data: Two case studies. *Insights the UKSG journal*, 32:1, 1–11. <http://doi.org/10.1629/uksg.463>
- Julkaisufoorumi. <https://www.julkaisufoorumi.fi/fi>
- Jytilä, Riitta & Mikael Laakso (2019) *Selvitys avoimesta vertaisarvioinnista kotimaisen tiedejulkaisemisen kentällä*. Tieteellisten seurain valtuuskunnan verkkojulkaisu 5. <https://doi.org/10.23847/978-952-599516-9>
- Kansallinen julkaisumetriikkaopas (2022) <https://wiki.eduuni.fi/display/CSCbibliometriikka/Kansallinen+julkaisumetriikkaopas>
- Karvonen, Erkki & Terttu Kortelainen & Jarmo Saarti (2014) *Julkaise tai tuhoudu! Johdatus tieteelliseen viestintään*. Tampere: Vastapaino.
- Koerber, Amy & Jesse Starkey & Karin Ardon-Dryer & R. Glenn Cummins & Lyombe Eko & Kerk F. Kee (2020) A qualitative content analysis of watchlists vs safelists: How do they address the issue of predatory publishing? *The journal of academic librarianship*, 46:6, 102236. <https://doi.org/10.1016/j.acalib.2020.102236>
- Kortelainen, Terttu (2010) Tieteellinen julkaisutoiminta ja informetrinen tutkimus. Teoksessa Sami Serola (toim.) *Ote informaatiosta. Johdatus informaatiotutkimukseen ja interaktiiviseen mediaan*. Helsinki: BTJ Kustannus, 244–263.

- Kulczycki, Emanuel & Raf Guns & Janne Pölonen & Tim C. E. Engels & Ewa A. Rozkosz & Alesia A. Zuccala & Kasper Bruun & Olli Eskola & Andreja Istenič Starčič & Michal Petr & Gunnar Sivertsen (2020) Multilingual publishing in the social sciences and humanities: A seven-country European study. *Journal of the association for information science and technology*, 71:11, 1371–1385. <https://doi.org/10.1002/asi.24336>
- Kyvik, Svein (1991) *Productivity in academia: Scientific publishing in Norwegian universities*. Oslo: Universitetsforlaget
- Laakso, Mikael (2021) *Open at any price? Reviewing short- and long-term implications of current forms of open access*. <https://avointiede.fi/en/news/cost-convenience-need-steering-wheel-open-access-shopping-cart>
- Late, Elina & Laura Korkeamäki & Janne Pölonen & Sami Syrjämäki (2020) The role of learned societies in national scholarly publishing. *Learned publishing*, 33:1, 5–13. <https://doi.org/10.1002/leap.1270>
- Late, Elina & Carol Tenopir & Sanna Talja & Lisa Christian (2019) Reading practices in scholarly work: from articles and books to blogs. *Journal of documentation*, 75:3, 478–499. <https://doi.org/10.1108/JD-11-2018-0178>
- Larivière, Vincent & Stefanie Haustein & Philippe Mongeon (2015) The oligopoly of academic publishers in the digital era. *Plos one*, 10:6, e0127502. <https://doi.org/10.1371/journal.pone.0127502>
- Martín-Martín, Alberto & Mike Thelwall & Enrique Orduna-Malea & Emilio López-Cózar (2021) Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics* 126, 871–906. <https://doi.org/10.1007/s11192-020-03690-4>
- Meadows, Arthur Jack (1974) *Communication in science*. Lontoo: Butterworths.
- Mongeon, Philippe & Adèle Paul-Hus (2016) The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106:1, 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- OECD (2015) *Frascati manual 2015. Guidelines for collecting and reporting data on research and experimental development*. https://read.oecd-ilibrary.org/science-and-technology/frascati-manual-2015_9789264239012-en#page1
- OKM:n korkeakoulujen tiedonkeruukäsikirja (2022). <https://wiki.eduuni.fi/pages/viewpage.action?pageId=262279864>
- Panitch, Judith & Sarah Michalak (2005) *The serials crisis: A white paper for the UNC-Chapel Hill scholarly communications convocation*. https://ils.unc.edu/courses/2017_fall/inls700_001/Readings/Panitch2005-SerialsCrisis.htm
- Piro, Fredrik Niclas & Dag Aksnes & Kristoffer Rørstad (2013) A macro analysis of productivity differences across fields: Challenges in the measurement of scientific publishing. *Journal of the American society for information science and technology*, 64:2, 307–320. <https://doi.org/10.1002/asi.22746>
- Puuska, Hanna-Mari & Marita Miettinen (2008) *Julkaisukäytännöt eri tieteenaloilla*. Opetusministeriön julkaisu 2008:33. <http://urn.fi/URN:ISBN:978-952-485-566-2>
- Pölonen, Janne (2021) Kansalliset julkaisukanavat mahdollistavat tiedejulkaisemisen kotimaisilla kielillä. *Tietolinja* 2021:2. <https://urn.fi/URN:NBN:fi-fe2021121661234>

4.1 Tieteellinen julkaiseminen – mitä, miksi ja miten?

- Pölönen, Janne (2022) Julkaisufoorumi ja rahoitus- ja arviointijärjestelmien muutokset Pohjoismaissa ja Euroopassa. *Tieteessä tapahtuu* 4/2022. <https://www.tieteessatapahtuu.fi/numerot/4-2022/julkaisufoorumi-ja-rahoitus-ja-arviointijarjestelmien-muutokset-pohjoismaissa-ja>
- Pölönen, Janne (2023) Julkaisufoorumi ja yliopistojen rahoitusmalli. *Tieteessä tapahtuu* 2/2023. <https://www.tieteessatapahtuu.fi/numerot/2-2023/julkaisufoorumi-ja-yliopistojen-rahoitusmalli>
- Rantala, Jukka (2012) Julkaisufoorumi-hanke huolestuttaa humanisteja ja yhteiskuntatieteilijöitä. *Kasvatus & aika* 6:1, 3–5.
- Ross-Hellauer, Tony (2017) What is open peer review? A systematic review. *F1000Research*, 6, 588–30. <http://doi.org/10.12688/f1000research.11369.1>
- Rousi, Antti & Mikael Laakso (2020) Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research. *Scientometrics*, 124:1, 131–152. <https://doi.org/10.1007/s11192-020-03467-9>
- San Francisco declaration of research assessment (2010) <https://sfdora.org>
- Sivertsen, Gunnar (2016) Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics* 107, 357–368. <https://doi.org/10.1007/s11192-016-1845-1>
- Sivertsen, Gunnar (2019) Understanding and evaluating research and scholarly publishing in the social sciences and humanities (SSH). *Data and information management*, 3:2, 61–71. <https://doi.org/10.2478/dim-2019-0008>
- Spezi, Valerie & Simon Wakeling & Stephen Pinfield & Claire Creaser & Jenny Fry & Peter Willett (2017) Open-access mega-journals: The future of scholarly communication or academic dumping ground? A review. *Journal of documentation*, 73:2, 263–283. <https://doi.org/10.1108/JD-06-2016-0082>
- Stoy, Lennart & Rita Morais & Lidia Borrell-Damián (2019) *Decrypting the big deal landscape. Follow-up of the 2019 EUA big deals survey report*. <https://eua.eu/resources/publications/889:decrypting-the-big-deal-landscape.html>
- Talja, Sanna & Carol Tenopir & Elina Late (2021) Desired affordances of scholarly e-articles: Views from scholars based on open-ended answers. *Libri*, 72:1, 67–81. <https://doi.org/10.1515/libri-2021-0012>
- Tennant, Jonathan & Harry Crane & Tom Crick & Jacinto Davila & Asura Enkhbayar & Johanna Havemann & Bianca Kramer & Ryan Martin & Paola Masuzzo & Andy Nobes & Curt Rice & Bárbara Rivera-López & Tony Ross-Hellauer & Susanne Sattler & Paul Thacker & Marc Vanholsbeeck (2019) Ten hot topics around scholarly publishing. *Publications*, 7:2, 34. <https://doi.org/10.3390/publications7020034>
- Tiedejatutkimus.fi. <https://tiedejatutkimus.fi/>
- Tutkimuseettinen neuvottelukunta (2012) *Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsittelyminen Suomessa*. https://tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf
- Unesco (2021) *UNESCO Recommendation on open science*. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>
- van Dalen, Hendrik & Kène Henkens (2012) Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. *Journal of the American society for information science and technology*, 63:7, 1282–1293. <https://doi.org/10.1002/asi.22636>

- Vastuullisen tutkijanarvioinnin työryhmä (2020) *Tutkijanarvioinnin hyvät käytännöt. Vastuullisen tutkijanarvioinnin kansallinen suositus*. Vastuullisen tieteen julkaisusarja 5:2020. <https://doi.org/10.23847/isbn.9789525995268>
- Weller, Ann C. (2001) *Editorial peer review: Its strengths and weaknesses*. Information Today, Inc.
- Wilsdon, James & Liz Allen & Eleonora Belfiore & Philip Campbell & Stephen Curry & Steven Hill & Richard Jones & Roger Kain & Simon Kerridge & Mike Thelwall & Jane Tinkler & Ian Viney & Paul Wouters & Jude Hill & Ben Johnson (2015) *The metric tide. Report of the independent review of the role of metrics in research assessment and management*. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Ylijoki, Oili-Helena & Anu Lyytinen & Liisa Marttila (2011) Different research markets: a disciplinary perspective. *Higher education* 62, 721–740. <https://doi.org/10.1007/s10734-011-9414-2>

4.2 Sähköinen asiakirjahallinta hakee vielä uomiaan

Pekka Henttonen

- Asiakirjatiedon hallinta on 1990-luvulta alkaen perustunut sähköisiin asiakirjajärjestelmiin, jotka varmistavat tiedon ymmärrettävyyden, autenttisuuden, eheyden ja luotettavuuden sekä elinkaaren ja käyttöoikeuksien systemaattisen hallinnan.
- Suomessa Kansallisarkiston SÄHKE-määritys (2005, 2008) kuvaa näiden järjestelmien vaatimukset. Suurin osa tietojärjestelmistä on kuitenkin tietokantoja ja muita kuin asiakirjajärjestelmiä.
- Koko elinkaaren digitalisoituminen on kesken. Kansallisarkisto on vasta siirtymässä kokonaan sähköiseen arkistointiin. Valtaosa arkistoaaineistoista on analogisessa muodossa. Samalla kun uudet aineistot ovat syntysähköisiä, analogisia paperiaineistoja kuitenkin hitaasti digitoidaan.
- Sähköisten aineistojen käyttöön liittyy kuitenkin suuria ongelmia, jos käyttörajoitettuja tai hävitettäviä osia tiedosta ei voida tunnistaa. Tietoa on liikaa, jotta sitä voitaisiin manuaalisesti käydä läpi, eikä esimerkiksi tekoäly ole ratkaisuna riittävän luotettava.

Arkistotiede eli asiakirjahallinnan tutkimus on perinteisesti sijoittunut eri opetus- ja tutkimusalueiden risteyskohtaan. Coxin¹ mukaan alueen ensimmäisen sukupolven tutkijat väittelivät Yhdysvalloissa yleensä historiatieteessä ja toisen sukupolven tutkijat kirjastotieteessä. 1990-luvulta lähtien asiakirjahallinnan tutkimusta ja ope-

1 Cox ym. 2001.

tusta on ollut yhä enemmän myös itsenäisenä tutkimusalueena.² Suomessa alueen yliopisto-opetus alkoi ensimmäisenä Tampereen yliopistossa vuonna 1997 osana informaatiotutkimusta.³

Klassinen arkistotiede tarkasteli asiakirjoja ja arkistoja fyysisinä objekteina.⁴ *Asiakirjan* määrittelemisen ei ole yksinkertaista, koska käsitettä voidaan lähestyä eri näkökulmista. Yksi raja-asiakirjalle saadaan lainsäädännöstä, mutta on hyödyllisempää ajatella asiakirjoja tietona, joka syntyy organisaatioiden ja yksityisten henkilöiden hoitaessa tehtäviään. Tätä tietoa säilytetään toiminnan tukemiseksi ja yhteiskunnan laajempiin tarkoituksiin, kuten kulttuurihistoriallista tutkimusta varten.⁵

Sähköistymisen myötä asiakirjoilta ja arkistoilta katosi niiden fyysinen muoto. Nyt sähköisistä aineistoista on jo vuosikymmenten kokemus. Käsitteet niiden ongelmista ovat muuttuneet ajan myötä. Alkuun suurimpana ongelmana pidettiin tallenteiden lyhyttä fyysistä ikää, esimerkiksi magneettinauhat eivät kestä kovin pitkään hapertumatta. Tallenteiden fyysinen ikä on parhaimmillaankin yleensä vain vuosikymmeniä eikä samaa luokkaa vuosisatoja säilyvän hyvälaatuisen paperin kanssa. Tämä on kuitenkin vain yksi puoli asiasta. Laitteet ja ohjelmistot vanhenevat, joten vanhaa dataa ei välttämättä pysty enää lukemaan ja muuttamaan ihmisen ymmärtämään muotoon, vaikka se sinänsä olisikin säilynyt. Kolmas kerros tulee merkitysten säilyttämisestä: ei auta, että data on säilynyt käytettävissä, jos ei tiedetä, mitä se merkitsee. Jos esimerkiksi sukupuoli on kyselytutkimuksen vastauksissa koodattu ykkösellä ja kakkosella, mutta on unohdettu merkitä ylös koodien merkitys, tieto on menetetty. Tämä ongelma torjutaan riittävällä dokumentaatiolla.⁶

Sähköiset aineistot olivat aluksi yleensä rekistereitä ja tutkimusdataa. Varsinaiset asiakirja-aineistot säilytettiin mikrofilmi- tai pape-

2 Ks. esim. Poole & Todd-Diaz 2022.

3 Valtonen 1999.

4 Henttonen 2015.

5 Henttonen 2023a.

6 Ks. esim. Henttonen 1999.

ritulosteina silloinkin, kun tieto oli sähköistä. 1990-luvulle tultaessa tämä toimintalinja ei enää toiminut. Oikeuden päätöksissä todettiin asiakirjaksi nimenomaan sähköinen asiakirja, ei sen tuloste. Asiakirjahallinnassa oli ryhdyttävä miettimään, mitä tämä tarkoittaa. Tätä selvitettiin tutkimusprojekteissa Brittiläisen Kolumbian ja Pittsburghin yliopistoissa.⁷

Projektit nostivat esille tarpeen säilyttää asiakirjojen väliset yhteydet ja niiden yhteyden toimintaan.⁸ Projektien tuloksia hyödynnettiin 1990-luvun lopulla, jolloin sähköisille asiakirjajärjestelmille alettiin julkaista vaatimusmäärittelyjä muun muassa Yhdysvalloissa, Kanadassa ja Uudessa-Seelannissa.⁹ Asiakirjajärjestelmien vaatimusmäärittelyiden tarkoitus on varmistaa, että asiakirja-aineistoa voi käyttää todisteena toiminnasta ja että se on ymmärrettävää ja hallittavaa. Vaatimukset tyypillisesti edellyttävät, että asiakirjajärjestelmässä on tehtäviä kuvaava luokitus, johon asiakirjat linkitetään. Kun asiakirja tallennetaan järjestelmään, siihen lisätään metatietoa kuvaamaan sen yhteyksiä muihin asiakirjoihin. Kun asiakirja on valmis, se lukitaan niin, ettei sitä pysty enää muuttamaan tai poistamaan järjestelmästä ilman tarvittavia valtuuksia. Hallintametatiedoilla estetään asiaton pääsyn asiakirjoihin. Niiden perusteella tietoja voidaan myös poistaa, kun asiakirjan säilytysaika on ummessa.¹⁰

Suomessa kehitys on seurannut kansainvälisiä kulkuja. Ensimmäiset ohjeet tietokonetallenteiden arkistoinnista annettiin vuonna 1987.¹¹ Ajalle ominaiseen tyyliin niissä kiinnitettiin huomiota tallennusvälineiden laatuun ja aineiston dokumentaation. Ohjeessa todettiin, että ”yleensä tiedostoista määrätään pysyvään säilytykseen paperi- tai mikrofilmituloste”. Tämä olikin lähes aina säilytysratkaisu. Kansallisarkisto julkaisi ensimmäisen version omasta kansallisesta asiakirjajärjestelmävaatimusmäärittelystä (SÄHKE) vuonna 2005.

7 Marsden 1997.

8 Lappin ym. 2021.

9 Gable 2002; Wilhelm, 2009.

10 Bantin 2008.

11 Valtionarkisto 1987.

Tämän jälkeen tuli mahdolliseksi hakea arkistolaitokselta lupaa tietojen säilyttämiselle pelkästään sähköisesti. Edellytyksenä oli järjestelmän SÄHKE-yhteensopivuus.¹² Ensimmäisen luvan sähköiseen säilyttämiseen sai Rautatievirasto vuonna 2008.¹³

Samalla kun asiakirjajärjestelmät saivat huomiota, myös muuta kehitystä tapahtui. Standardointi eteni suurin askelin. Asiakirjahallinnon kansainvälinen ISO 15489 -standardi ilmestyi vuonna 2001. Siinä kuvataan toiminta, jota sähköiset asiakirjajärjestelmät tukevat. Standardi määrittelee, millaisia piirteitä asiakirjatiedolla on: se on autenttista, ehyttä, luotettavaa ja käytettävää. *Autenttisuus* tarkoittaa sitä, että tieto on ”sitä, mitä se sanoo olevansa”. Lisäksi tiedetään, kuka on tiedon laatinut tai lähettänyt ja millä hetkellä. Tieto on *ehyttä*, jos sitä ei ole tahattomasti tai tahallisesti muutettu, ja *luotettavaa*, jos sen voidaan uskoa olevan täydellinen esitys niistä seikoista, mistä se todistaa. *Käytettävä* tieto on löydettävissä ja esitettävissä. ISO 15489 kuvaa myös yleisesti sen, mitä hyvä asiakirjahallinta organisaatiossa tarkoittaa.¹⁴

Arkistoteoriassa merkittävä askel oli niin sanotun elinkaaren *jatkumomallin* kehittäminen. Perinteisesti asiakirjatiedon elinkaarta oli hahmotettu lineaarisesti vain eteenpäin menevinä vaiheina, joissa käyttäjät ja aineistosta vastaavat tahot vaihtuivat: aineistot ovat ensin viranomaisella käytössä (aktiivivaihe), sitten niitä käytetään yhä harvemmin (passiivivaihe), ja lopulta toiminnassa tarpeettomat asiakirjat, jos niitä ei hävitetä, siirretään säilytettäväksi Kansallisarkistoon tai muuhun päätearkistoon (historiallinen vaihe). Tämä perinteinen elinkaarimalli soveltuu erityisen huonosti sähköisille aineistoille, koska viimeisen päätearkistovaiheen tarpeet on huomioitava elinkaareissa alusta alkaen. Vaihtoehdoksi muotoiltiin elinkaaren jatkumomalli, jossa aineiston ajatellaan olevan ainakin potentiaalisesti merkittävää vuoroin tai samanaikaisesti monesta eri

12 Helenius 2011.

13 Rautatievirasto ensimmäisenä sähköiseen arkistointiin 2008 <https://agricolaverkko.fi/keskustelu/viewtopic.php?t=1737> (haettu 12.8.2008).

14 SFS-ISO 15489-1 2017.

näkökulmasta, esimerkiksi sekä kulttuurihistoriallisesti että organisaation toiminnan kannalta.¹⁵

Suuria odotuksia 2000-luvun alussa

Kehitystä siis tapahtui. 2000-luvun vaihteessa sähköisiin asiakirjajärjestelmiin kohdistui paljon odotuksia. Niiden uskottiin ratkaisevan sähköisen asiakirjahallinnan ongelmat. Samalla visioitiin ammattikunnan roolin muutosta: arkistonhoitaja nousisi kellarin arkistoholvista johtoryhmien pöytään asettamaan vaatimuksia tietojärjestelmille.

Nyt voi nähdä, että näin ei käynyt. Sähköiset asiakirjajärjestelmät ovat asiakirjahallinnan näkökulmasta toimivia, ja organisaatiot, joissa niitä käytetään, ovat paremmassa asemassa kuin muut, koska pääsyä tietoihin ja tiedon elinkaarta voidaan hallita näissä järjestelmissä siten kuin lainsäädäntö edellyttää. Asiakirjat myös toimivat todisteina siitä, miten on toimittu. Asiakirjajärjestelmistä ei ole kuitenkaan tullut sellaista tiedonhallinnan kokonaisratkaisua kuin toivottiin. Asiakirjajärjestelmät ovat usein vaikeasti käytettäviä, ja niiden integrointi muihin työssä käytettyihin tietojärjestelmiin saattaa olla huono. Sähköiset asiakirjajärjestelmät ovat vain yksi organisaatioissa käytetyistä järjestelmätyypeistä ja osa digitaalisen tiedon kasvavaa kirjoa. Niinpä arkistonhoitaja on palannut kellariin, joka nyt elektronisena on entistä enemmän sekaisin.¹⁶

Tällä hetkellä sähköisten asiakirjajärjestelmien vaatimusmäärittelyjen kehitys on monin paikoin pysähtynyt. Eurooppalaisesta MoRegista ilmestyi nähtävästi viimeinen versio (MoReq2010) vuonna 2011. Norjassa on päätetty lopettaa kansallisen NOARK-määrittelyn kehittäminen.¹⁷ Suomessa Kansallisarkisto kumosi SÄHKE-määrittelyn aseman määräyksenä vuoden 2022 lopussa. SÄHKE on edel-

¹⁵ Malleista, ks. An 2003.

¹⁶ Lappin 2010.

¹⁷ Arkivverket 2021.

leen olemassa suosituksena.¹⁸ Tämä voi merkitä asiakirjajärjestelmien vastatuulta, mutta myös sitä, että tietty tietojärjestelmien kypsyyssaso on jo saavutettu eikä enää vaadita yhtä määrätietoista ohjaamista ja kehittämistä kuin ennen. Tähän viitataan Norjassa tehdyssä päätöksessä: voimavarat suunnataan nyt enemmän standardointia vaativille alueille. Tällainen on esimerkiksi menetelmä, jolla tiedon autenttisuudesta voitaisiin varmistua.¹⁹ Myös Suomessa viimeisimmän arkistolakiehdotuksen perusteluissa todetaan optimistisesti SÄHKE-määritysten olevan keskeisin osin jo huomioituja tietojärjestelmissä.²⁰

Suunta näyttää olevan nyt kohti ratkaisuja, joissa arkistoinnin vaatimukset otetaan huomioon yleisesti tietojärjestelmien suunnittelussa.²¹ Samalla painopiste on siirtynyt enemmän tiedon käytettävyyteen tiedon eheyden, luotettavuuden ja autenttisuuden takamisesta.²²

Onko tieto pitkäaikaisesti säilytettävää?

Sähköisessä pitkäaikaissäilytyksessä on kaksi keskeistä tekijää. Ensimmäinen on se, millaista tietoa voidaan parhaiten säilyttää. Toinen on se, millaista tietoa viranomaiset tuottavat. Ihanteellisesti nämä kohtaisivat täydellisesti – viranomaisten pitkäaikaisesti arvokas tieto on myös säilytykseen sopivaa – mutta käytännössä näin ei aina ole. Kehityksessä voikin nähdä yrityksiä sovittaa niitä yhteen paremmin.

Millaista aineistoa voidaan säilyttää?

Kuten jo todettiin, sähköisten aineistojen ongelmaksi paljastuivat ensin tallennusvälineiden hauraus ja lyhytikäisyys ja sitten tallenteiden ohjelmisto- ja laitteistoriippuvuus. Ratkaisuksi on kehitetty

18 Kansallisarkisto 2022.

19 Arkivverket 2021.

20 Opetus- ja kulttuuriministeriö 2021, 27.

21 Ns. Archiving by Design.

22 Henttonen 2024b.

erilaisia strategioita, joilla esimerkiksi vahingoittuneilta tallenteilta pelastetaan tietoa (digitaalinen arkeologia). Vanhentunutta laitteisto- ja ohjelmistoympäristöä voidaan myös pitää edelleen yllä ("tietokonemuseo") tai sen voidaan antaa kadota ja mennyttä ympäristöä emuloida ohjelmallisesti. Tavallisin strategia on valita joi-takin standardoituja, todennäköisesti hyvin pitkäikäisiä ja laajasti käytettäviä tiedostomuotoja, joissa tieto tallennetaan pitkäaikaissäilytysjärjestelmään pidettäväksi siinä yllä. Tällaisia muotoja voivat olla esimerkiksi PDF/A-dokumentti tai TIFF-kuvatiedosto. Pitkäaikaissäilytyksen näkökulmasta hyvä tieto on sellaista, että se on jo alkujaan sopivassa muodossa tai ainakin muunnettavissa siihen ilman, että tietoa katoaa. Pitkäaikaissäilytys on lisäksi sitä yksinkertaisempaa ja tehokkaampaa, mitä homogeenisempaa aineisto on. Kun tallennettujen tiedostotyyppien, tietorakenteiden ja metatietomallien määrä on vähäinen, säilytyksen suunnittelu, ylläpito ja riskit yksinkertaistuvat verrattuna tilanteeseen, jossa säilytyksessä on lukuisia räätälöityjä toimenpiteitä vaativia aineistoryhmiä.

Pelkkä datan säilyminen ei kuitenkaan usein riitä, vaan tiedon pitäisi olla autenttista, ehyttä, luotettavaa ja käytettävää. SÄHKE ja muut asiakirjajärjestelmien vaatimusmäärittelyt ovat vastaus tähän ongelmaan. Asiakirjajärjestelmien toiminnoilla ja metatiedoilla varmistetaan, että syntyvä tieto täyttää asiakirjatiedon vaatimukset.

Myös aineistonhallinta asettaa vaatimuksia aineistolle. Kustannustehokas säilytys kohdistuu vain sellaiseen tietoon, joka on todella tarpeen säilyttää. Toisaalta voi olla myös tietoa, jota ei saisi säilyttää juridisista syistä. Tällainen tieto on voitava erottaa muusta ja hävittää. Lisäksi on tietoa, joka on yksilön tietosuojan vuoksi tai muusta syystä käyttörajoituksen alaista. Tämänkin tiedon on oltava tunnistettavissa. Sekä elinkaaren että käyttörajoitusten hallinta on huomioitu sähköisissä asiakirjajärjestelmissä.

Päätearkiston intressissä on asettaa säilytykseen tulevalle aineistolle sekä teknisiä että sisällöllisiä vaatimuksia, jotta asiakirjatiedon laatu sekä aineiston käytettävyys ja ylläpidettävyys ovat varmistettuja pitkäaikaissäilytyksessä.

Pitkä tie sähköiseen elinkaareen

Laadukas ja helposti säilytettävä sähköinen aineisto vaatii hyvin suunniteltua aineistohallintaa, mikä taas edellyttää osaamista ja väivännäköä. Viranomaisen asiakirjahallinnon on oltava riittävän korkealla tasolla, mutta tuottavatko viranomaiset säilytyksen kannalta hyvää asiakirjatietoa?

Cunningham²³ arvioi tilannetta australialaisesta näkökulmasta. Australian kansallisarkisto päätti vuonna 1994 poikkeuksellisesti jättää aineistot säilytettäväksi viranomaisten omiin järjestelmiin. Yksiy oli halu välttää kalliit sähköisen säilyttämisen kehittämisprojektit ja oppia muiden tekemisistä.²⁴ Maan kansallisarkistolla ei ollut säilyttämiseen riittävästi resursseja, ja tällä tavoin vältettiin myös ongelmat, jotka seuraavat tietojen säilytyksestä alkuperäisen laitteisto- ja ohjelmistoympäristön ulkopuolella.

Ajatus oli aluksi, että aineistot olisivat jatkossa hajallaan eri viranomaisten järjestelmissä kansallisarkiston asettaessa vaatimuksia säilyttämislle ja toimiessa verkoston keskuspiirteenä, jonka kautta eri tahoilla olevat aineistot löydetään. Tätä *postkustodialista* visiota (termi viittasi siihen, että aineistot eivät olleet arkiston hallussa) arvosteltiin tuolloin. Arkistoteoreetikot kritisoiivat kansallisarkistoa säilytysvastuun väistelystä. Viranomaisille arkistointi ei ollut niiden päämäärätoimintaa, joten ne eivät halunneet siitä koituvia kustannuksia ja olivat innokkaita sysäämään aineistonsa kansallisarkistolle.²⁵

Australiassa strateginen valinta oli kuitenkin kehittää ensin viranomaisten sähköistä asiakirjahallintaa ja vasta sitten pitkäaikais-säilytystä. Katsottiin, ettei ollut mielekäästä keskittyä pitkäaikais-säilytykseen ennen kuin viranomaisilta olisi säilytykseen saatavissa kunnollisia asiakirja-aineistoja. Australialaisen kehitystyön tulok-sena syntyi monia kansainvälisesti keskeisiä asiakirjahallinnan standardeja ja malleja, kuten kansainvälinen asiakirjahallinnon ISO

23 Cunningham 2011.

24 Cunningham 2008.

25 Cunningham 2011.

15489 standardi, menetelmä toimintaprosessien kuvaamiseksi sekä *Designing and implementing recordkeeping systems* (DIRKS) -menetelmä asiakirjahallinnan suunnitteluun. Silti, kun Australian kansallisarkisto oli viimein saanut pitkäaikaissäilytysjärjestelmänsä valmiiksi, ei vuonna 2011 ollutkaan viranomaisia, jotka olisivat siirtäneet siihen aineistojaan. Mikä meni pieleen?²⁶

Digitaalisen datan puutteesta viranomaisissa ei ainakaan ole kyse. Datan määrä on kasvanut eksponentiaalisesti. Samalla myös säilytyskustannukset ovat räjähtäneet, vaikka säilytyksen yksikkökustannus väheneekin koko ajan. Ongelma on, että viranomaiset eivät pysty erottamaan säilytettävää hävitettävistä. Seulonta on niin vaivalloista, että on helpompaa säilyttää kaikki. Vain kolmannes Australian viranomaisista oli kattavasti siirtynyt sähköiseen asiakirjahallintaan ja 40 prosenttia ei ollut edes suunnitellut sitä.²⁷ Cunningham sanookin, että vaikka standardeja on kehitetty ja ohjeistusta laadittu, on selvästi eri asia saada ne läpi viranomaisissa. Hän arvostelee Australian kansallisarkistoa siitä, että se unohti tavoitteet ja keskittyi toimintatapoihin. Informaatiota ei käytetä pelkästään informaation vuoksi, vaan yleensä se liittyy johonkin muuhun ihmisten toimintoon tai tehtävään.²⁸ Uudet ohjeet edellyttivät tehtäväpohjaista asiakirjahallintaa, mikä edellytti tehtävien analyysiä ja asiakirjahallinnan vaatimusten määrittämistä, mikä oli kaikille uutta. Kansallisarkisto peitteli epävarmuuttaan keskittymällä jäykkään ja monimutkaiseen suunnitteluprosessiin. Vaikka asiakirjahallinnan suunnittelumenetelmänä DIRKS on periaatteessa selkeä ja joustava, vähäiset resurssit menivät viranomaisissa asiakirjaprosessien uudelleen keksimiseen ja työn päämäärä unohtui. Tämä teki DIRKSistä kirosanan ja DIRKS-projekteista loputtomia ja tarkoituksettomia.²⁹

²⁶ Cunningham 2011.

²⁷ Cunningham 2011.

²⁸ Kumpulainen & Sihvonon 2025, tämän kirjan luku 2.3 Tehtävälähtöinen informaatiovuorovaikutus.

²⁹ Cunningham 2011.

Onko Suomessa onnistuttu paremmin? Ei välttämättä. Näyttää siltä, että myös täällä ollaan kaukana ihannetilanteesta, jossa tiedonhallinta olisi täysin digitaalista, seulottava aineisto erotettavissa muusta, sähköisen tiedon elinkaari aina suunniteltu, säilytysaikapäätökset ajantasaisia ja kattavia, elinkaarta hallittaisiin systemaattisesti ja metatiedot aineistossa olisivat riittäviä ja yhdenmukaisia sekä hallinnan että myöhemmän käytön tarpeisiin. Tällöin aineiston siirto viranomaiselta päätearkistoon olisi periaatteessa yksinkertaista, ja se voitaisiin tehdä vain todella säilytettävän aineiston osalta. Tällöin myös tietopalvelu voitaisiin helposti toteuttaa lainsäädännön asettamien oikeuksien ja rajoitusten mukaisesti.

Todellisuus on toinen. Vaikka julkishallinto saattaa mielikuvissa vaikuttaa digitalisoituneelta, täysin sähköiset prosessit ja tiedon elinkaari ovat meilläkin harvinaisia. Vain noin yksi kymmenestä valtionhallinnon toimijasta sanoi vuonna 2020, että pysyvästi säilytettävien tai arkistoitavien asiakirjojen elinkaari on kokonaan digitaalinen ja prosessit sähköisiä alusta loppuun.³⁰ Samaan aikaan yhdessäkään yliopistossa ei ollut pysyvästi säilytettävälle aineistoille täysin digitaalista elinkaarta.³¹ Kunnissakin tiedonhallinta oli yleensä osin hybridiä.³² Siirtyminen täysin digitaaliseen tiedonhallintaan on viranomaisissa siis vielä kesken.

Suomalaisen asiakirjahallinnan kehittämiseksi viranomaisissa on tehty työtä. Resursseja käytettiin 2010-luvun vaihteessa ensinnäkin yhteisten tehtäväluokitusten tekemiseen. Siinä ei nähtävästi ole kuitenkaan kaikilta osin päästy alkuperäisiin tavoitteisiin. Yhteisten tehtäväluokitusten laatimisessa ongelmia ovat aiheuttaneet muun muassa monitulkintaiset käsitteet ja epäselvyys työn tavoitteista, tehtäväluokitusten erilaiset kontekstit ja käyttöympäristöt, työhön osallistuvien vaihtelevat panostukset ja tahtotilat sekä asiakirjahal-

30 Mäkiranta 2020.

31 Kokkinen 2020.

32 Hänninen 2020.

linnan resurssien yleinen puutteellisuus, mikä tekee suurten uudistusten läpiviemisestä työlästä.³³

Toinen kehityspanostus on ollut niin sanotun *tiedonohjauksen* edellyttämien tiedonohjaussuunnitelmien laatiminen ja tiedonohjauksen toteuttaminen viranomaisten tietojärjestelmissä. Tiedonohjaus tuli käsitteistöön SÄHKE2-määrityksen myötä vuonna 2008. Sen ideana on, että erillisessä tiedonohjausjärjestelmässä oleva tiedonohjaussuunnitelma tuottaa asiakirjajärjestelmän lisäksi muihinkin tietojärjestelmiin metatietoja ja ohjaa asiakirjahallinnan prosesseja esimerkiksi asettamalla tiedolle säilytysaikoja ja päivittämällä käyttöoikeustietoja prosessin edetessä hallinnossa. Paradigmaattinen esimerkki tästä on se, miten valmisteluasiakirjat tulevat yleensä julkisiksi, kun asiassa on tehty päätös. Vanhemmissa *arkistonmuodostussuunnitelmissa* listattiin asiakirjojen säilytysajat, käyttörajotukset ja hallinnan kannalta keskeiset tiedot, kuten säilytyspaikat, asiakirjojen keskinäinen järjestys ja rekisteröintitarve. *Tiedonohjaussuunnitelmissa* on lisäksi oltava aineistoon liittyvien prosessivaiheiden kuvaus, jotta tiedonohjaus mahdollistuu.

Tiedonohjaukseen ei näytä täyttäneen siihen asetettuja toiveita. Selvitykset viittaavat siihen, että se on toteutettu yleensä vain asiakäsittelyjärjestelmissä ja järjestelmissä, jotka palvelevat vastaavia hallintoprosesseja. Sellaiset järjestelmät ovat toki hallinnossa yleisiä, mutta ne eivät ole läheskään ainoa järjestelmätyyppi. Tiedonohjauksen visiona oli ohjata erilaisia tietojärjestelmiä laajemmin,³⁴ mutta muut toiminnassa käytetyt niin sanotut *operatiiviset tietojärjestelmät*, vaikkapa Puolustusvoimien tilannekuvan seurantajärjestelmä tai yliopiston oppimisympäristö, saattavat olla niin kaukana tiedonohjauksen edellyttämästä mallista, ettei tiedonohjauksen toteuttaminen niissä olisi mahdollista ainakaan ilman erityistapauksiin sopivia joustavia määrittelyjä.³⁵ Suurin osa julkishallinnon toiminnassa käytetyistä tietojärjestelmissä on tiedonohjauksen ulko-

33 Henttonen 2018.

34 Henttonen 2023b.

35 Hänninen 2020.

puolella. Toimiva tiedonohjaus vaatisi tarkkaa suunnittelua, mutta suunnittelun taso viranomaisissa vaihtelee. Jonkinlainen tiedonohjaussuunnitelma (tai vanhempi arkistonmuodostussuunnitelma) löytyy kyllä valtion viranomaisilta laajasti, mutta sen toteutustapa ja sisältö vaihtelevat.³⁶ Pienissä kunnissa ei ole ehkä lainkaan suunnitelmaa. Joskus suunnitelmassa on säilytysajat ja -tavat, muttei tiedonohjauksen edellyttämiä prosessitietoja. Jos prosessivaiheetkin ovat suunnitelmassa, niin voi olla, ettei suunnitelmaa ole integroitu tietojärjestelmiin, jolloin tiedonohjaus ei toteudu.³⁷

Tämä saattaa olla yksi syy myös kansallisen pitkäaikaissäilytysjärjestelmän etenemiseen takkuilen. Maaliskuussa 2011 otettiin käyttöön Kansallisarkiston sähköinen säilytysjärjestelmä, vastaanotto- ja palvelujärjestelmä VAPA. Organisaatioiden todellinen valmius hyödyntää palvelua arvioitiin kuitenkin heikoksi muun muassa järjestelmistä puuttuvien siirtomahdollisuuksien vuoksi.³⁸ Valtion viranomaiset vaikuttivat haluttomilta tekemään itsenäisiä päätöksiä sähköisen asiakirjahallinnon kehittämisestä, ja sähköisen säilyttämisen lupia haettiin harvaksen.³⁹ Vain viitisentoista sähköisen säilyttämisen hakemusta oli vuonna 2011 joko hyväksytty tai käsiteltävänä.⁴⁰ Kansallisarkiston kyselyyn vastanneista viranomaisista vain alle puolet suunnitteli palvelun käyttöönottoa.⁴¹ VAPAan oli vaikeuksia saada viranomaisilta sähköisiä aineistoja, ja palvelu ajettiin lopulta muutaman vuoden jälkeen alas vuonna 2015.⁴²

Tulos oli siis samantapainen kuin Australiassa, mutta syyt ehkä osin toiset. Ollakka kollegoineen⁴³ on todennut, että sähköisen arkistoinnin ongelmaksi on nähty opetus- ja kulttuuriministeriön yleisluontoinen ja tehoton ohjaus, arkistolaitoksen huonot resurssit

36 Mäkiranta 2020.

37 Hänninen 2020.

38 VAPA avattu asiakkaiden käyttöön, Akti 1/2011.

39 Eerikäinen 2011.

40 Helenius 2011.

41 Ahtiainen 2011.

42 Karkimo 2015.

43 Ollakka ym. 2014.

ja vakiintuneen ohjausasetaman suhde vähäisempiin muodollisiin valtuuksiin, valtiovarainministeriön määrittelemätön rooli ohjauksessa sekä sähköisen asiakirjahallinnan epäselvät ohjausvastuut yleensä. Päätearkiston kannalta riittävä ja sen ydintehtävää hoitava palvelu ei tarjonnut viranomaisille riittävästi kannustimia käyttää sitä. VAPA ei vastannut viranomaisten tarpeisiin, koska siihen saattoi luovuttaa vain pysyvästi säilytettävää tietoa. Viranomaisen oli joka tapauksessa järjestettävä määrääjän säilytettävän aineiston säilytys – joka saattaa olla hyvin pitkäaikaista sekin – ja huolehdittava tietopalvelusta.⁴⁴ Kyse on joukosta yhteen liittyviä kysymyksiä: kuka omistaa aineiston, kuka hallinnoi sitä ja hoitaa tietopalvelua, kenen käyttöä varten aineistoa säilytetään, mikä on palveluun otettavan aineiston säilytysaika ja miten kustannukset jaetaan.

Suuremmalla haavilla aineistoa arkistoon

Päätearkiston intressissä on vaatia aineistolta laatua. Kansallisarkiston normi- ja informaatio-ohjauksen tavoitteena on, että säilytettävään tietoon voidaan luottaa.⁴⁵ Toimia pidetään tarpeellisena sekä asiakirjojen validiteetin että tieteellisen tutkimuksen tarpeiden vuoksi.⁴⁶ Päätearkiston näkökulmasta on tärkeää, että sen saama aineisto on paitsi laadukasta (ymmärrettävää, käytettävää, ehyttä ja luotettavaa) myös tasalaatuista, eli se sisältää vain pysyvästi säilytettäviä asiakirjoja, jolloin sitä ei tarvitse enää seuloa. Lisäksi erilaista hallintaa vaativien osien olisi oltava kauttaaltaan tunnistettu, mikä tarkoittaa ennen kaikkea käyttörajoitusten alaisen tiedonaineksen erottamista muista. Onnistunut ja tehokas seulonta sekä käyttörajoitusten hallinta palvelevat myös viranomaisen tiedonhallintaa. Asiakirjahallinnon suunnittelu ja tiedonohjaus ovat keinoja, joilla tähän tavoitteeseen pyritään.

44 Ollakka ym. 2014.

45 Happonen 2020.

46 Nuorteva 2020.

Muitakin keinoja voidaan käyttää. Jos erottelu on vaikeaa tai työstä, voidaan poistaa tarve erotella. Jos esimerkiksi pysyvästi ja eri määräaikoja säilytettävän tiedon erottelu on ongelma, sitä voidaan kiertää tekemällä kategorisointia karkeammalla tasolla. Tästä on se etu, että kun eri säilytysaikakategorioita on vähän, tiedon luokittelun osumatarkkuus paranee, varsinkin jos tavalliset käyttäjät joutuvat siitä huolehtimaan.⁴⁷ Suomessa Kansallisarkisto on koikeillut seulptastrategiassa laajojen kategorioiden strategiaa toisella tavalla eli määräämällä säilytettäväksi sekä kokonaisia tehtäväluokkia asiantkäsittelyjärjestelmissä – sen sijaan että eri asiakirjatyypeille (kuten päätös, muistio tai saapunut kirje) määriteltäisiin tehtäväluokan sisällä eri säilytysaikoja – että kokonaisia tietojärjestelmiä ilman, että niiden tietosisällöstä eroteltaisiin säilytettäviä osia.⁴⁸ Lähestymistapa ei tietävästi ole aina ollut toimiva. Viranomaiset ovat vastustaneet suurempien aineistokokonaisuuksien säilyttämisestä koituvia kuluja. Toinen ongelma on, että lainsäädäntö edellyttää henkilötietojen säilytyksen minimoimista ja tietojen hävittämistä sen jälkeen, kun niitä ei enää toiminnassa tarvita. Tästä näkökulmasta suurien, myös henkilötietoja säilyttävien kokonaisuuksien säilyttäminen sellaisenaan on ongelmallista. Henkilötietojen säilytys näyttää olevan yksi seulonnan kipupisteistä, koska historiallista aineistoa säilyttävän päätearkiston näkökulmasta myös yksilötason tietojen säilyttäminen on usein tarpeen.⁴⁹

Varmasti laadukasta asiakirjatietoa saadaan asiantkäsittelyjärjestelmistä, mutta ne ovat vain yksi tietojärjestelmätyyppi muiden joukossa, eikä tiedonohjauksella ole saatu muita järjestelmiä asiakirjahallinnon haltuun. Vaikka SÄHKE2-määräystä kuvailtiin tietojärjestelmäriippumattomaksi, Kansallisarkistossa ymmärrettiin, että se oli ensisijaisesti sovellettavissa dokumenttipohjaisiin järjestelmiin. Tämä jätti määräyksen ulkopuolelle tietorakenteeltaan toisenlaisia järjestelmiä, kuten tietokantoja, rekistereitä ja

47 Cisco 2008.

48 Eräkaski 2014; Leppänen 2016.

49 Eräkaski 2014; Orrman 2012; Rasimus 2019.

tutkimusaineistoja. Näiden Kansallisarkistoon siirtämistä varten suunniteltiin niin sanottua SÄHKE₃-määritystä, jonka toivottiin astuvan voimaan vuosien 2013–2014 aikana.⁵⁰ Näin ei käynyt, mutta tietokantojen säilytys uudessa säilytyspalvelussa (SAPA) perustuu jo silloin tarkasteltavana olleeseen SIARDiin.⁵¹ Siinä tietokannan sisällöstä luodaan XML-muotoinen säilytyspaketti, josta tietokanta voidaan tarvittaessa luoda uudelleen.⁵² Tämä laajentaa kirjoa, jota sähköiseen säilytysjärjestelmään voidaan ottaa.

Harppauksin sähköiseen arkistointiin

Kun Kansallisarkiston aikaisempi vastaanotto- ja säilytyspalvelu VAPA ajettiin alas, aineistojen vastaanotossa oli muutaman vuoden katkos, kunnes se alkoi uudestaan vuoden 2018 alussa.⁵³ Julkishallinnon uusi sähköisen arkistoinnin palvelu (SAPA) perustuu arkistojen, kirjastojen ja museoiden yhteiselle KP-PAS-palvelulle.⁵⁴ Vuodesta 2025 palvelu on koko julkishallinnon käytössä.⁵⁵

Analogisesta paperimaailmasta siirrytään kohti sähköisyyttä. Aikaisemmasta mallista, jossa sähköinen pysyvä säilyttäminen oli luvan vaativa poikkeus, on siirrytty päivastaiseen käytäntöön. Kansallisarkisto ottaa nyt vastaan viranomaisilta pääsääntöisesti vain digitaalista aineistoa.⁵⁶ Vuoden 2022 alusta aineistot on lähtökohtaisesti arkistoitu digitaalisina, vaikka ne eivät olisikaan syntysähköisiä. Analogiset, esimerkiksi paperiset, asiakirjat digitoidaan ja sen jälkeen hävitetään.⁵⁷ Niin sanotussa massadigitointiprojektissa digitoidaan viranomaisilla olevat paperiaineistot vuoteen 2030 mennessä.⁵⁸ Digitoiduissa aineistoissa ei ole metatietoja, mutta niihin lisätään

50 Eräkaski 2011a; Eräkaski 2011b.

51 Etelävuori 2020.

52 SIARDista tarkemmin, ks. Eriksson 2017.

53 Merenmies 2017.

54 Opetus- ja kulttuuriministeriö 2021.

55 Kansallisarkisto 2024.

56 Kansallisarkisto 2020.

57 Kansallisarkisto 2021a.

58 Eräkaski 2017.

kuvailutiedot, kun ne siirretään Kansallisarkistolle.⁵⁹ Analogisena säilyttämistä tehdään vain, jos alkuperäisellä asiakirjalla katsotaan olevan erityistä arvoa analogisena.⁶⁰ Digitointi on Ukrainan sodan varjossa alettu nähdä keinona suojata kulttuurihistoriallisesti merkittäviä aineistoja poikkeusoloissa.⁶¹

Australian tavoin Suomessakaan ei ole paperiasiakirjojen vastaanottamiseen enää tiloja eikä tilojen lisärakentamiseen halukkuutta. Myös Yhdysvaltain kansallisarkisto NARA on tehnyt vastaavan päätöksen ja siirtynyt digitaalisen aineiston vastaanottoon. Tämä siirtyminen kohti tiedon elinkaaren täyttä sähköistymistä on ainakin poliittinen, ja kenties arkistojenkin, tahtotila. Sipilän hallituksen periaatepäätöksessä linjattiin vuonna 2017 sekä paperiaineistojen digitoinnista että siirtymisestä sähköiseen arkistointiin.⁶²

Samalla kun asiakirjasiirrot viranomaisilta Kansallisarkistoon muuttuvat sähköisiksi, arkistossa jo olevia kokoelmia digitoidaan – hitaasti. Vuonna 2009 Kansallisarkiston digiarkistossa oli noin 1,5 miljoonaa asiakirjaa.⁶³ Pelkästään vuonna 2021 arkistossa olevia asiakirjoja digitoitiin takautuvasti moninkertainen määrä, yhteensä 9,3 miljoonan tiedoston verran. Digitoituja tiedostoja on kaikkiaan jo 87 miljoonaa. Se on kuitenkin vain neljä prosenttia aineistoista.⁶⁴ Tätä vauhtia koko Kansallisarkistossa olevan aineiston digitointi kestää 200 vuotta.⁶⁵ Toisaalta 80 prosenttia käyttötarpeesta saataisiin katettua digitoimalla kaikkein käytetyimmät 20 prosenttia aineistoista.⁶⁶ Kansallisarkisto on tehnyt digitointia myös yhteistyössä sellaisten ulkopuolisten tahojen kanssa, joilla on intressi aineistojen saattamiseen käyttöön digitaalisessa muodossa. Esimerkiksi Suo-

59 Merenmies ym. 2018.

60 Kansallisarkisto 2019.

61 Verma 2022.

62 Opetus- ja kulttuuriministeriö 2017.

63 Kecskeméti 2009.

64 Kansallisarkisto 2021b.

65 Rosvall 2022.

66 Kansallisarkisto 2021b.

men sodissa vuosina 1939–1945 kaatuneiden sotilaskantakortit digitoitiin yhteistyössä sukututkimuspalvelu Ancestryn kanssa.⁶⁷

Tällä hetkellä (2025), sähköisen asiakirjahallinnan kenttä näyttää täysin toiselta kuin vajaa vuosikymmen sitten. Aikaisemmin Kansallisarkisto oli sähköisen asiakirjahallinnan kehittäjä ja määräysten antaja.⁶⁸ Tämä on nyt muuttunut. Tiedonhallintalakia⁶⁹ valmisteltaessa arkistointi rajattiin lain ulkopuolelle, koska Kansallisarkiston asemasta ei päästy yksimielisyyteen,⁷⁰ ja arkistolain uudistaminen päätettiin tehdä erikseen. Tuloksena on, että yleisohjaus jää valtiovarainministeriölle ja sen yhteydessä toimivalle, valtioneuvoston asettamalle tiedonhallintalautakunnalle.⁷¹ Tiedonhallintalautakunta johtaa tiedonhallintaa (ja siis myös asiakirjahallintaa – mitä termiä ei käytetä tiedonhallintalaissa) ohjeistamalla tiedonhallintalain täytäntöönpanoa viranomaisissa ja valvomalla viranomaisten toimintaa lain osalta. Kansallisarkiston rooli on rajoittumassa vain päätearkistotoimintaan. Tiedonohjaus ollaan sikäli unohtamassa, että uusimman arkistolakiehdotuksen mukaan ”SÄHKE2-määräys ei tule enää ohjaamaan tulevaisuudessa tiedonohjausta”, vaikka tietojärjestelmien määrittelyä ja metatietojen keräämistä pidetään edelleen tärkeänä. Samassa lakiehdotuksessa kunnille luodaan mahdollisuuksia käyttää Kansallisarkiston keskitettyä säilytyspalvelua.⁷² SÄHKE2 kumottiin 31.12.2022. Se on edelleen olemassa suosituksena. Kansallisarkiston verkkosivulla SÄHKE2-mukaisuutta luonnehditaan ”tuetuksi tietorakenteeksi Kansallisarkiston sähköisen arkistoinnin palveluun”.⁷³

Täsmällisesti Kansallisarkiston asemaa tiedonhallinnan kentässä ei ole kuitenkaan ratkaistu, koska vuonna 1994 säädetyn arkistolain

67 KecsKeméti 2018.

68 Henttonen 2024a.

69 Laki julkisen hallinnon tiedonhallinnasta 906/2019, 2019.

70 Nuorteva 2020.

71 Nuorteva 2020.

72 Opetus- ja kulttuuriministeriö 2021.

73 Ks. <https://kansallisarkisto.fi/usein-kysyttyja-kysymyksiä> (haettu 25.8.2023).

uudistaminen on edelleen tekemättä. Sitä on yritetty vuodesta 2013 alkaen jo viiden eri hallituksen voimin.⁷⁴

Säilytettävien sähköisten aineistojen kirjo kasvaa

Perinteisesti säilytykseen on pyritty ottamaan asiakirjahallinnan näkökulmasta laadukkaita ja helposti säilytettäviä sähköisiä aineistoja, mutta koska kaikki siirrot Kansallisarkistoon ovat sähköisiä, näyttää väistämättömältä, että kriteerejä digitaaliselle aineistolle joudutaan höllentämään. Aikaisemmin Kansallisarkisto otti vastaan sähköisiä aineistoja SÄHKE-yhteensopivista asianhallintajärjestelmistä, mutta tästä ei voida pitää enää kiinni. Tällä hetkellä vaatimuksia asetetaan tiedostotyypeille ja niin sanotun siirtopaketin muodostamiselle.⁷⁵

Tästä herää kysymyksiä. Vaatimukset eivät riittäne ”asiakirjallisuuden” takaamiseen sellaisena kuin se asiakirjajärjestelmien määrittelyssä on nähty. Säilytettävän tiedon autenttisuutta, luotettavuutta, eheyttä ja ymmärrettävyyttä ei toisin sanoen ole varmistettu samoin kuin asiakirjajärjestelmissä, joten sähköiseen pitkäaikaissäilytykseen tulee tietoa, joka ei välttämättä täytä asiakirjatiedon kriteereitä ainakaan siten kuten ne ymmärrettiin 2000-luvun vaihteen vaatimusmäärittelyissä asiakirjajärjestelmille. On epäselvää, millainen merkitys tällä lopulta on.

Selvemmät haasteet liittyvät siihen, että vastaanotettavat aineistot ovat todennäköisesti puutteellisesti seuloituja eli ne sisältävät myös tietoja, jotka olisi pitänyt hävittää, ja ettei niistä ole tunnistettavissa käyttörajoituksiltaan erilaisia osia. Australiassa on todettu, että kun ei pystytä erottamaan tärkeää vähemmän tärkeästä, säilytetään kaikki.⁷⁶ Suomessa viranomaisille vuonna 2014 tehdyssä kyselyssä vain yhdeksän prosenttia vastaajista kertoi tietoja hävitetyn tietojärjestelmistä, kun määritelty säilytysaika oli päättynyt. Vajaa

74 Voutilainen 2023.

75 Kansallisarkisto 2018; 2021c.

76 Cunningham 2011.

kolmannes kertoi, että näin oli tapahtunut satunnaisesti. Kuusi kymmenestä vastaajasta sanoi, ettei tietoja ole lainkaan hävitetty tai säilytysaika on määrittelemättä.⁷⁷ Kuntien tiedonohjaussuunnitelmat eivät ole kattavia, joskus sellaisia ei edes ole.⁷⁸ Suunnitelmien kattavuus ja ajantasaisuus on yksi kysymys, mutta selvittämättä on, missä määrin säilytysaikapäätökset ovat toteutettavissa muissa kuin SÄHKEn mukaisissa asiankäsittelyjärjestelmissä. Sama koskee käyttöoikeuksien hallintaa ja erilaisten käyttörajoitusten alaisten osien tunnistettavuutta tiedoissa.

Sähköisen pääsyn ongelma

Kysymykset ovat erityisen tärkeitä, koska digitaalisen informaation määrä päätearkistoissa tulee moninkertaistumaan. Baron⁷⁹ arvioi, että tämän vuosisadan puolivälissä, siis kolmenkymmenen vuoden kuluttua, Yhdysvaltain kansallisarkisto NARAN aineistot ovat suurimmaksi osaksi ja ehkä valtaosin digitaalisia. Kysymys on, miten tähän aineistomassaan voidaan tarjota pääsy. Ongelmana ovat sekä aineiston seulonta, tarvittavan tiedon löytäminen että sensitiivisten tietojen tunnistaminen. Tietoa on liikaa, jotta sitä voitaisiin käydä käsin läpi. Ongelma on tiedossa ja siihen on yritetty kehittää sekä luonnollisen kielen teknologiaan (NLP) että tekoälyyn pohjautuvia ratkaisuja,⁸⁰ mutta tulokset eivät ole täydellisiä. Parhaaseen tulokseen päästään ihmisen ja algoritmien yhteistyöllä.⁸¹ Työkalujen olisi oltava iteratiivisia, jolloin asiantuntija arvioi algoritmin tuottamia tuloksia ja korjaa siinä olevia virheitä, jolloin tekoäly oppii entistä paremmaksi.⁸² Digitaalisen tiedon lupaus on kuitenkin välitön pääsy tietoon verkon yli ilman välikäsiä.

77 Lampi 2015.

78 Hänninen 2020.

79 Baron 2022.

80 Hutchinson 2017; 2020; Lee 2018.

81 Baron 2022.

82 Hutchinson 2020.

Ongelmat eivät ole ennen kuulumattomia. Myös analogiseen aineistoon pääsy on voinut olla vaikeaa. Vaikka monet oikeudelliset ja kiinteistöjä koskevat asiakirjat ovat olleet Yhdysvalloissa periaatteessa saatavilla, niihin ei ole ollut käytännössä pääsyä, koska tiedon käsittelykustannukset ovat olleet liian korkeat.⁸³ Uutisissa on kerrottu tapauksesta, jossa tutkija pyysi Kanadassa käyttöönsä 20 vuotta vanhan korruptiotutkinnan aineistoa. Kanadan kansallisarasto yritti torjua pyynnön todettuaan aineiston läpikäynnin vaativan 96 vuotta ja maksavan 21 miljoonaa Kanadan dollaria. Kyse oli paperi- ja mikrofilmiaineistosta, jota oli 780 000 sivua.⁸⁴

Digitaalisen aineiston kohdalla ongelmiin törmätään kuitenkin uudessa mittakaavassa. Sähköpostit tarjoavat esimerkin siitä, mikä voi olla tulos. Yhdysvalloissa presidenttikauden päättyessä presidenttihallinnon aineistot siirretään lain nojalla NARAan. NARAssa oli jo vuonna 2017 säilytyksessä 750 miljoonaa sähköpostiviestiä presidenttien ja liittovaltion hallinnosta 1980-luvun alusta Reaganin kaudesta alkaen. Se on enemmän kuin kaikkien presidenttien kirjastoissa on painettua materiaalia.⁸⁵ Valkoisen talon ja korkeimman oikeuden 535 miljoonasta viestistä kuitenkin vain 0,08 prosenttia on avattu yleisön käyttöön.⁸⁶ Koska datasta ei voida tunnistaa helposti ja nopeasti käyttäjän siitä tarvitsemaa osaa, johon tämä laillisesti voisi päästä, koko aineisto pysyy käytännössä suljettuna.

Käykö Suomessa samoin arkistojen täysin sähköistyessä? Käytössä olevien tietojen perusteella se näyttää mahdolliselta. Toisaalta tilanteessa on tuntemattomia tekijöitä, joiden vaikutusta ei vielä tunneta. Yksi niistä on lainsäädäntö. Jos sähköistyminen johtaisi arkistojen sulkeutumiseen, se olisi erityisen suuri muutos Pohjoismaissa, joissa hallinnon ja päätöksenteon läpinäkyvyydessä keskeistä on kansalaisten pääsy viranomaistietoon.⁸⁷ Toisaalta nimen-

83 Perritt & Rustad 2000.

84 Bronskill 2022.

85 Borden & Baron 2016.

86 Baron 2022.

87 Grønbech-Jensen 1998.

omaan pohjoismainen hallintokulttuuri voi tässä tilanteessa olla toimivampi, koska julkisuuslainsäädäntöön liittyvä päätöksenteko on periaatteessa yksinkertaista ja julkisuus aina lähtökohta.⁸⁸ Samalla täälläkään ei voitane välttää tilanteita, joissa salassa pidettävä tieto on erotettava muusta. Tähän yhdysvaltalainen malli saattaa sopia paremmin. Yhdysvaltain vuodesta 1966 voimassa olleen Freedom of Information lainsäädännön ajatus on, että kansalainen voi pakottaa viranomaisen julkistamaan tietoa, joka ei muuten olisi saatavilla. Tämä on toiminut esimerkkinä monille maille niiden laatiessa omaa julkisuuslainsäädäntöään.⁸⁹ Olennaista on, että yhdysvaltalaiseen aineistoon pääsyyn lähtökohtaisesti sisältyy enemmän viranomaisen harkintaa kuin meillä Suomessa. Jos käyttörajoituksen alaista aineistoa ei kuitenkaan pystytä luotettavasti tunnistamaan kohtuullisessa ajassa, molemmat mallit ovat vaikeuksissa.

Niinpä voi olettaa, että tuleva kehitys on arkistojen digitalisoiduessa kahtalainen. Samalla kun uudemmat syntysähköiset aineistot eivät mahdollisesti tule lainkaan yleisesti saataville edellä kuvattujen vaikeuksien takia, vanhemmat analogiset aineistot saavat digitoituina tietoverkkojen kautta yhä laajemman käyttäjäkunnan. Arkistot ovat edelleen vasta uuden ajan kynnyksellä, eivätkä koko elinkaaren sähköistymisen vaikutukset ole vielä selvinneet.

88 Ks. Voutilainen 2025, tämän kirjan luku 4.3 Oikeustieteen ja informaatiotutkimuksen yhteyksiä etsimässä.

89 Perritt & Rustad 2000.

Lähteet

- Ahtiainen, Eeva-Kaisa (2011) Ministeriö selvisi VAPA-pilotoinnista. *Akti* 1/2011, 6–7.
- An, Xiaomi (2003) An integrated approach to records management. *The information management journal* 37:4, 24–30
- Arkivverket (2021) Arkivverket avslutter utviklingen av Noark-standarden. <https://www.arkivverket.no/nyheter/arkivverket-avslutter-utviklingen-av-noark-standarden-copy> (haettu 6.8.2021).
- Bantin, Philip C. (2008) *Understanding data and information systems for recordkeeping*. Lontoo: Facet.
- Baron, Jason R. (2022) Challenges in providing access to the digital universe. *Aeolian network workshop # 3*. <https://www.aeolian-network.net/workshop-3-prof-jason-r-baron-challenges-in-providing-access-to-the-digital-universe-are-algorithms-the-answer/>
- Borden, Bennett B & Jason R. Baron (2016) Opening up dark digital archives through the use of analytics to identify sensitive content. *2016 IEEE international conference on big data (big data)*, 3224–3229. <https://doi.org/10.1109/BigData.2016.7840978>
- Bronskill, Jim (2022) Man seeking historical records told it would cost feds \$21 million to process request. *Toronto Star* 6.3.2022. <https://www.thestar.com/politics/2022/06/03/man-seeking-historical-records-told-it-would-cost-feds-21-million-to-process-request.html>
- Cisco, Susan (2008) How to win the compliance battle using 'big buckets'. *Information management journal* 42:4, 30–38.
- Cox, Richard J. & Elizabeth Yakel & David Wallace & Jennifer Marshall (2001) Archival education in North American library and information science schools. *The library quarterly* 71:2, 141–194. <https://doi.org/10.1086/603260>
- Cunningham, Adrian (2008) Digital curation/digital archiving: A view from the national archives of Australia. *The American archivist* 71:2, 530–543. <https://doi.org/10.17723/aarc.71.2.pohot68547385507>
- Cunningham, Adrian (2011) Good digital records don't just "happen": embedding digital recordkeeping as an organic component of business processes and systems. *Archivaria* 71, nro Spring, 21–34.
- Eerikäinen, Nina (2011) Valtion viranomaisen VAPA- ja VALDA-viidakossa. *Akti* 2/2011, 17.
- Eriksson, Heidi (2017) Relaatietietokantojen pitkäaikaissäilytys XML-konversion avulla systemaattinen kirjallisuuskatsaus. Pro gradu, Tampereen yliopisto. <https://urn.fi/URN:NBN:fi:uta-201710122601>
- Eräkaski, Mikko (2011a) SÄHKE3 luo edellytykset tietokantojen ja rekisterien pysyvälle säilytykselle. *Akti* 2/2011, 16.
- Eräkaski, Mikko (2011b) Miten säilyttää tietokantojen tiedot. *Akti* 3/2011, 16.
- Eräkaski, Mikko (2014) Ennakkoseulonnan merkitys korostuu sähköisessä toimintaympäristössä. *Akti* 3/2014, 15.
- Eräkaski, Mikko (2017) Massadigitoinnin tiekartta. *Akti* 2/2017, 6–7.
- Etelävuori, Reko (2020) SAPA-alustan toteutussuunnitelma 2020–2022. Esitetty tilaisuudessa SAPA-info viranomaisille, 12.11.2020, Kansallisarkisto, Helsinki.

4.2 Sähköinen asiakirjahallinta hakee vielä uomiaan

- Gable, Julie (2002) Everything you wanted to know about DoD 5015.2. *The information management journal* 36:6, 33–38.
- Grønbech-Jensen, Carsten (1998) The Scandinavian tradition of open government and the European Union: Problems of compatibility? *Journal of European public policy* 5:1, 185–199. <https://doi.org/10.1080/1350176880000091>
- Happonen, Päivi (2020) Asiakirjahallinnan neuvonnan ja ohjauksen kysyntä suurempi kuin koskaan. *Akti* 1/2020, 3.
- Helenius, Armi (2011) Kokemuksia sähköisen säilyttämisen lupaprosessista. *Akti* 3/2011, 17.
- Henttonen, Pekka (1999) Atk-aineistojen arkistoinnin haasteita. *Arkistoyhdistyksen julkaisuja* 6/1999, 23–64.
- Henttonen, Pekka (2015) *Johdatus asiakirjahallinnan tutkimukseen*. Helsinki: Avain.
- Henttonen, Pekka (2018) Creating common functional classifications. Experiences of the Finnish public-sector records management. Teoksessa Aida Škoro Babić (toim.) *Archives in the service of people. People in the service of archives*. Alma Mater Europea, 93–99.
- Henttonen, Pekka (2023a) *Johdatus asiakirjahallinnan tutkimukseen*. Uudistettu laitos. Helsinki: Avain.
- Henttonen, Pekka (2023b) One system to rule them all” – or how to manage records across information systems. Teoksessa Marianne Rostgaard ja Greg Bak (toim.) *The Nordic model of digital archiving*. New York: Routledge <https://doi.org/10.4324/9781003325406-9>.
- Henttonen, Pekka (2024a) Archives and public agencies in Finland. A relationship in transformation. *Atlanti* 34:1, 41–55.
- Henttonen, Pekka (2024b) Mitä asiakirjajärjestelmien jälkeen? *Faali* 1/2024, 20.
- Hutchinson, Tim (2017) Protecting privacy in the archives. Preliminary explorations of topic modeling for born-digital collections. *2017 IEEE international conference on big data (big data)*. Boston: IEEE, 2251–2255. <https://doi.org/10.1109/BigData.2017.8258177>.
- Hutchinson, Tim (2020) Natural language processing and machine learning as practical toolsets for archival processing. *Records management journal* 30:2, 155–174. <https://doi.org/10.1108/RMJ-09-2019-0055>.
- Hänninen, Heli (2020) Sähköisen arkistoinnin tilannekuvan selvitys kunnan toimialoilla [Julkaissamaton]. Valtiovarainministeriö, Suomen Kuntaliitto ry.
- Kansallisarkisto (2018) *Digitaaliseen säilytykseen hyväksyttävät tiedostoformaattit*. Helsinki: Kansallisarkisto.
- Kansallisarkisto (2019) *Kansallisarkiston arvottamisen kategoriat ja kriteerit analogisten asiakirjojen säilytysmuodosta päättämiseksi. Täydentävä liite seulostrategiaan*. KA/3742/07.01.01.03.00/2019. Helsinki: Kansallisarkisto.
- Kansallisarkisto (2020) *Analogisten asiakirjojen luovuttaminen ja siirrot*. Helsinki: Kansallisarkisto.
- Kansallisarkisto (2021a) *Määräys arkistoitavien asiakirjojen muodosta*. KA/15906/07.01.01.00/2021. Helsinki: Kansallisarkisto.
- Kansallisarkisto (2021b) *Toiminnan ja talouden suunnitelma 2023–2026*. KA/14096/00.01.00/2021”, Helsinki: Kansallisarkisto.

- Kansallisarkisto (2021c) *Vastaanotettavia tiedostoja koskevat vaatimukset 1.0*. Helsinki: Kansallisarkisto.
- Kansallisarkisto (2022) Määräys Sähke2-määräyksen kumoamisesta KA/12266/07.01.01.00/2022. Helsinki: Kansallisarkisto.
- Kansallisarkisto (2024) Sähköisen arkistoinnin palvelusta kansallinen ratkaisu. *Kansallisarkiston uutiskirje viranomaisille*, 4/2024. <http://uutiskirje.arkisto.fi/some/html/2080534>
- Karkimo, Ari (2015) Valtio tallensi kiintolevyllisen tietoja miljoonilla euroilla – nyt se loppuu. *Tietoviikko* 26.8.2015.
- Kecskeméti, István (2009) Kansallisarkisto sai jättipotin digitointiin. *Akti* 2/2009, 6.
- Kecskeméti, István (2018) Takautuvaan digitointiin tarvitaan uusia toimintatapoja. *Akti* 2/2018, 6–7.
- Kokkinen, Susanna (2020) Sähköisen arkistoinnin nykytilan selvitys. Korkeakoulusektori. Aalto-yliopisto, 6.12.2020.
- Laki julkisen hallinnon tiedonhallinnasta 906/2019 (2019). <http://www.finlex.fi/fi/laki/alkup/2019/20190906>
- Lampi, Mervi (2015) Hyvään tiedonhallintaan on vielä matkaa. *Akti* 1/2015, 10–11.
- Lappin, James (2010) What will be the next records management orthodoxy? *Records management journal* 20:3, 252–264. <https://doi.org/10.1108/09565691011095283>.
- Lappin, James & Tom Jackson & Graham Matthews & Claire Ravenwood (2021) Rival records management models in an era of partial automation. *Archival science* 21, 243–266. <https://doi.org/10.1007/s10502-020-09354-9>.
- Lee, Christopher A. (2018) Computer-assisted appraisal and selection of archival materials. 2018 *IEEE international conference on big data (big data)*, 2721–2724. Seattle: IEEE. <https://doi.org/10.1109/BigData.2018.8622267>.
- Leppänen, Markku (2016) Ennätysmäärä seulontapäätöksiä. *Akti* 1/2016, 9.
- Marsden, Paul (1997) When is the future? Comparative notes on the electronic record-keeping projects of the University of Pittsburgh and the University of British Columbia. *Archivaria* 43, 158–73.
- Merenmies, Markus (2017) Kansallisarkiston digikolmiloikka tuottaa uusia palveluja. *Akti* 3, 18.
- Merenmies, Markus & Kenneth Ahlfors & Miia Herrala & Riia Viitanen (2018) AHAA tulossa käyttöön. *Akti* 1, 8.
- Mäkiranta, Minna-Liisa (2020) Sähköisen arkistoinnin nykytilan selvitys / valtionhallinto. Helsinki: Kansallisarkisto.
- Nuorteva, Jussi (2020) Kohti vuotta 2025 - pitääkö Kansallisarkiston uusi strategia? *Akti* 1/2020, 6–8.
- Ollakka, Esko & Pirkko Lahdelma & Eveliina Vainio (2014) *Sähköisen arkistoinnin edistäminen. Tuloksellisuustarkastuskertomus 11/2014*. Helsinki: Valtiontalouden tarkastusvirasto. <http://urn.fi/urn:isbn:978-952-499-272-5>
- Opetus- ja kulttuuriministeriö (2017) Valtioneuvoston periaatepäätös asiakirjallisen aineiston digitoinnista ja arkistoinnista vain sähköisenä (liite 2). Helsinki: Opetus- ja kulttuuriministeriö
- Opetus- ja kulttuuriministeriö (2021) *Luonnos hallituksen esitykseksi eduskunnalle arkistolain ja kansallisarkistosta annetun lain muuttamisesta*. Helsinki: Opetus- ja kulttuuriministeriö.

4.2 Sähköinen asiakirjahallinta hakee vielä uomiaan

- Orrman, Eljas (2012) Asiakirjojen seulonasta ja julkisuudesta tutkimuksen edellytyksiä säätelevinä tekijöinä. *Tieteessä tapahtuu* 30:3, 26–33.
- Perritt, Henry H. & Zachary Rustad (2000) Freedom of information spreads to Europe. *Government information quarterly* 17:4, 403–417. [https://doi.org/10.1016/S0740-624X\(00\)00050-2](https://doi.org/10.1016/S0740-624X(00)00050-2)
- Poole, Alex H., ja Ashley Todd-Diaz (2022) 'I'm not a very good visionary': Challenge and change in twenty-first century North American archival education. *Archival science* 22:4, 585–616. <https://doi.org/10.1007/s10502-022-09389-0>
- Rasimus, Tomi (2019) Julkisen hallinnon tietoaisteistojen säilyttäminen ja arkistointi. Pro gradu, Itä-Suomen yliopisto. <https://erepo.uef.fi/handle/123456789/22619>
- Rautatievirasto ensimmäisenä sähköiseen arkistointiin (2008). <https://agricolaverkko.fi/keskustelu/viewtopic.php?t=1737> (haettu 12.8.2008).
- Rosvall, Minna (2022) Digitaalinen kulttuuriperintö karttuu tuskaisen hitaasti – sota Ukrainassa herättää uusia huolia ainutkertaisten aineistojen katoamisesta. *Yle Uutiset* 23.3.2022. <https://yle.fi/uutiset/3-12368434>
- SFS-ISO 15489-1 (2017) Information and documentation. Records management. Part 1: Concepts and principles. Suomen Standardisointiliitto SFS ry.
- Valtionarkisto (1987) Arkistotoimen vaatimukset automaattisissa tietojenkäsittelyssä. *Valtionarkiston yleinen ohje nro 14*. Helsinki: Valtion painatuskeskus.
- Valtonen, Marjo Rita (1999) Records management research in Finland. *Records management journal* 9:3, 199–205.
- VAPA avattu asiakkaiden käyttöön (2011) *Akti* 1/2011.
- Verma, Pranishu (2022) Meet the 1,300 Librarians racing to back up Ukraine's digital archives. *Washington Post*, 4.8.2022. <https://www.washingtonpost.com/technology/2022/04/08/ukraine-digital-history/>
- Voutilainen, Tomi (2023) Arkistolain uudistaminen. Neljän vaalikauden ja viiden hallituksen hankeko? *Faali* 2/2023, 14–18.
- Wilhelm, Philipp (2009) An evaluation of MoReq2 in the context of national EDRMS standard developments in the UK and Europe. *Records management journal* 19, 117–133. <https://doi.org/10.1108/09565690910972075>

4.3 Oikeustieteen ja informaatiotutkimuksen yhteyksiä etsimässä

Tomi Voutilainen

- Artikkeliki tarkastelee oikeustieteen ja informaatiotutkimuksen välistä suhdetta tunnistamalla rajapintoja näiden väliltä.
- Oikeustieteellä on oma merkityksensä informaatiotutkimukselle, mutta myös informaatiotutkimuksella on läheinen suhde oikeustieteeseen.
- Aihetta käsitellään oikeusinformatiikan, oikeudellisen informaation tuottamisen, informaatio-oikeuden ja informaatiotutkimuksen tekemiseen vaikuttavan sääntelyn näkökulmista.

Informaatiotiedettä pidetään metatieteenä, johon sisältyy erilaisia informaation tutkimukseen kohdistuvia tutkimussuuntauksia.¹ Informaatiotieteen tarkoituksena on tuottaa tutkimustietoa tiedon hankintatavoista, käytöstä, organisoinnista, välittämisestä ja säilyttämisestä. Informaatiotieteen tavoitteena on tuottaa käsitteitä, menetelmiä ja järjestelmiä tiedon saatavuuden ja hyödynnettävyyden edistämiseksi.² Kansallisesti informaatiotutkimus rinnastuu sisällöltään pääosin informaatiotieteen tieteenalan määritelmään,³ joten tässä luvussa käytän pääasiassa informaatiotutkimuksen käsitettä.

1 Bates 1999, 1044.

2 Borko 1968, 3; Stock & Stock 2013, 3–4.

3 Mäkinen 2010, 11–12.

Oikeustiede on perusmuodossaan ja perusluonteeltaan informaation tuottamista, hallintaa ja tutkimista. Oikeustieteellisessä tutkimuksessa käsitellään oikeuslähteitä tai kerätään ja käsitellään erilaisiin oikeudellisiin ilmiöihin liittyviä tietoja empiirisiä menetelmiä käyttäen. Oikeustiede on myös informaatioympäristö, joka muodostuu erilaisista oikeuslähteistä, mutta oikeustiede tuottaa myös uusia oikeuslähteitä tutkimuksen avulla. Oikeustieteen keskeisenä tehtävänä on oikeudellisen informaation systematisointi ja tulkinta sekä voimassa olevan oikeuden sisällön esittäminen oikeuslähdeopin rakenteita hyödyntäen.⁴ Oikeustiede tieteenalana koostuu tutkimusaloista, joita ovat muun muassa oikeusdogmatiikka, oikeusteoria, oikeussosiologia, oikeushistoria, oikeustaloustiede, lainsäädäntötutkimus, oikeusinformatiikka ja oikeuslingvistiikka. Oikeustieteen tutkimus on siten sidoksissa monissa tilanteissa muihin tieteisiin kuten historiatieteeseen, sosiologiaan, taloustieteeseen, tietojenkäsittelytieteeseen ja informaatiotieteeseen.

Oikeustiede kohtaa informaatiotutkimuksen erityisesti kahdella tutkimusalalla – oikeuslingvistiikassa ja oikeusinformatiikassa. Oikeuslingvistiikka on tutkimusala, joka tutkii oikeuskieltä ja sen kehitystä, ominaisuuksia ja käyttöä sekä oikeuskielen termistöä ja sen syntaksia.⁵ Oikeuslingvistinen tutkimus on osa informaatiotutkimusta erityisesti silloin, kun tutkimuskohteena ovat oikeudelliset käsitteet ja niiden väliset suhteet taikka oikeudellisten aineistojen kuvaaminen oikeudellisilla käsitteillä. Toisaalta oikeudellisia käsitteitä kehitetään ja käytetään tulkinnan tukena myös oikeusdogmaattisessa tutkimuksessa toisinaan, jolloin puhutaan käsitelainopista, joka tosin on ollut pitkään oikeustieteissä kritiikin kohteena.⁶

Oikeusinformatiikka on puolestaan tutkimusala, jossa tutkitaan muun muassa oikeudellisen tiedon saatavuutta, oikeudellisia tietokantoja ja niiden käyttöä sekä oikeudellisia tietojärjestelmiä ja tiedonsiirtoa. Oikeusinformatiikan tutkimuksessa oikeutta pidetään

4 Aarnio 1989, 47–52 ja 220–221; Nuotio 2005, 130–131.

5 Mattila 2010, 719.

6 Hirvonen 2011, 45–46; Helin 2017, 896–900; Tuori 2002, 1301–1309.

informaatiojärjestelmänä.⁷ Oikeusinformatiikan tutkimuksessa käsitellään tyypillisesti oikeuden ja informaation sekä oikeuden ja tietoteknologian välisiä suhteita.⁸ Oikeusinformatiikan on katsottu olevan osa oikeustiedettä, joskin tutkimusalana se on monitieteellinen.⁹ Toisaalta oikeusinformatiikan käsitteessä informatiikka viittaa informaatiotieteisiin, joten sitä voidaan pitää myös informaatiotieteisiin kuuluvana tai ainakin läheisessä suhteessa siihen.¹⁰ Käsitteenä ja tutkimusalana *Legal Informatics* on katsottu osaksi informaatiotieteitä englanninkielisessä Wikipediassa, kun taas suomenkielisessä Wikipediassa oikeusinformatiikkaa ei ole sidottu niinkään informaatiotieteisiin, vaan enemmänkin teknologiasidonnaisiin tieteisiin.¹¹ Tämä selittyyne osaltaan sillä, että kansallisesti oikeusinformatiikan tutkimusalaa on hahmotettu ilman varsinaista teoreettista pohjaa tietoteknologian ilmiöihin painottuvana kuvailevana tutkimussuuntauksena. Kansallisessa tutkimusyhteisön haarassa tehty oikeusinformatiikan sisällön systematisointi ei ole myöskään saanut laajempaa jalansijaa kansainvälisessä oikeusinformatiikan tutkimuksessa.¹²

Oikeuslingvistiikka

Oikeustieteessä erityisesti oikeusdogmatiikassa käsitteillä on merkityksensä kunkin säädöksen sisältöä tulkittaessa. Eri säädökset muodostavat lähtökohtaisesti oman käsitteistönsä, joka määritellään säädöksessä. Kuitenkin sääntelystematiikan näkökulmasta käsitteistön hajautuminen ei ole toivottava ilmiö sääntelyn selkeyden kannalta katsottuna. Esimerkiksi julkisoikeudellisessa sääntelyssä monissa tilanteissa tiettyyn oikeudelliseen ongelmaan joudutaan

7 Siltala 2003, 112.

8 Saarenpää & Riekkinen 2023, 1.

9 Saarenpää 1999, 713–714.

10 Seipel 2005, 18.

11 Ks. englanninkielinen Wikipedia 2024a ja vertaa sitä suomenkieliseen Wikipediaan 2024b.

12 Ks. tästä jaottelusta esimerkiksi Saarenpää & Riekkinen 2023, 35–37.

hakemaan ratkaisua tulkitsemalla useita eri säädöksiä samanaikaisesti. Tällöin paitsi säädösten välisten suhteiden selkeydellä myös niiden käsiteharmonialla ja semanttisella yhteentoimivuudella on merkitystä sääntely-ympäristön ymmärrettävyyden ja tulkittavuuden kannalta katsottuna. Perustuslakivaliokunta onkin eri yhteyksissä kiinnittänyt huomiota siihen, että informaation käsittelyä koskeva sääntely on monin osin vaikeaselkoista ja että perusoikeuskysymyksen sääntelyn tulisi olla selkeää.¹³ Valiokunnan mukaan sääntelyn kohderyhmän tulee kyetä soveltamaan säännöksiä ilman vaikeuksia.¹⁴ Vaikeaselkoinen sääntely voi osoittautua tehottomaksi, koska tällaista sääntelyä on hankala noudattaa lainsäätäjän tahdon mukaisesti.¹⁵ Vaikeaselkoinen sääntely voi myös aiheuttaa oikeusturvaongelmia sääntelyn soveltajille.¹⁶ Lainvalmistelussa ei kuitenkaan ole käytössä varsinaisia käsitelmärittelyn oppeja, vaan käsitteet määritellään käytännössä lainvalmistelussa lainvalmistelijan työpöydällä.

Yleisesti ottaen viranomaisten on käytettävä hallintolain (434/2003) 9 §:n nojalla asiallista, selkeää ja ymmärrettävää kieltä. Lainkirjoittajan oppaan mukaan lakeja kirjoitettaessa käytetään lakikieltä, jonka pohjana on hyvä yleiskieli sekä asiallisuus, täsmällisyys ja yksitulkintaisuus.¹⁷ Vastaavasti oikeustieteellisessä kirjallisuudessa on esitetty yleisluonteisia kriteereitä hyvälle lakikielelle,¹⁸ kuten ymmärrettävyys, täsmällisyys ja käytetyn sanaston tuttuus. Lakikielen ohjeistus on teknisluonteista ja jättää lainvalmistelijalle varsin laajan liikkumavaran muun muassa käytettävän käsitteistön suhteen. Jossain määrin lainvalmistelutekniikassa hyödynnetään erilaisia fraaseja ja johtolauseita, jotka ovat suhteellisen vakioituja, mutta säädöksen käsitteistön muodostamiselle ei ole olemassa mi-

13 PeVL 4/2021 vp, 4, PeVL 31/2017 vp, 4 ja PeVL 45/2016 vp, 3.

14 PeVL 60/2014 vp, 3.

15 Voutilainen & Muukkonen 2021, 17.

16 PeVL 4/2021, 4.

17 Lainkirjoittajan opas 2013, 24.1

18 Kivivuori 1980, 9–17.

tään erityistä hyvää lainvalmistelutapaa ohjaavaa normistoa. Yleislakien kunnioittamisen periaate lainvalmistelussa kuitenkin ohjaa käsitteistön käyttöä siten, että erityislaeissa omaksutaan lähtökohdaisesti yleislain käsitteistö.¹⁹ Sinällään lainsäädännön kielenhuoltoa tehdään laintarkastuksessa ennen hallituksen esityksen antamista eduskunnalle, mutta se on varsin teknisluonteista.²⁰

Osaltaan kielen laatuun liitettävänä ongelmana on ollut Euroopan unionin sääntely, jossa säädöstekstit käännetään kaikille unionin virallisille kielille, myös suomeksi. Käännöstekstien kielellinen laatu ja käsitteistö eivät monissa tilanteissa vastaa laadultaan alkuperäistä säädöstekstiä, vaikka unionin virallisilla kielillä Euroopan unionin virallisessa lehdessä julkaistut versiot ovatkin todistusvoimaisia. Kielellisistä heikkouksista ja niiden korjaamisesta toimii esimerkkinä Euroopan unionin yleinen tietosuoja-asetus (EU) 2016/679, joka tuli voimaan toukokuussa 2016 ja jota alettiin soveltaa 25.5.2018 alkaen. Tietosuoja-asetuksen suomenkielinen versio sisälsi lukemattomia käännösvirheitä ja kielellisiä heikkouksia, mutta ne korjattiin vasta 4.3.2021 Euroopan unionin virallisessa lehdessä julkaistulla oikaisulla.²¹

Oikeudellisen kielen ja informaatiotutkimuksen kannalta katsottuna säädösten semantiikka ja käsitteiden väliset suhteet eivät ole lainvalmistelussa ja -säädännössä kelvollisella pohjalla. Toisaalta osin tutkimusperusteisesti on pyritty luomaan erilaisia käsite- tai termipankkeja myös oikeudellisista käsitteistä erityisesti julkiselle hallinnolle. Valitettavasti näihin termipankkeihin on luotu käsitteille merkityssisältöjä, jotka eivät vastaa laissa säädettyjen käsitteiden merkityssisältöjä. Esimerkiksi Suomalaiseen asiansanasto- ja ontologiapalveluun sisältyvät tietotermit poikkeavat varsin paljon lainsäädännössä määritellyistä käsitteiden sisällöistä, vaikka pohjana näyttääkin olevan lainsäädännössä omaksutut käsitteet ja nii-

19 Määttä 2015, 310–314.

20 Niemivuo 2020, 116.

21 Euroopan unionin virallinen lehti 2021.

den merkityssisällöt.²² Esimerkkinä tästä ongelmasta voi mainita asiakirjan ja henkilötiedon käsitteet. Tällainen lainsäädäntöön nähden rinnakkainen termien ja sanastojen kehittäminen on omiaan aiheuttamaan epäselvyyttä eri käsitteiden merkityssisällöstä eri alan asiantuntijoille. Erityisesti julkisessa hallinnossa sen toiminnan on kuitenkin nojaututtava perustuslain (731/1999) 2 §:n 3 momentin edellyttämällä tavalla lakiin, jota on tarkoin noudatettava. Muodostunut tilanne on jossain määrin absurdi, mutta myös osoittaa, että eri tieteiden välillä on jännitteitä, koska oikeustieteessä on totuttu luomaan oma käsitteympäristö ilman riittävää tieteidenvälistä vuoropuhelua. Toisaalta näiden ongelmien syihin voitaisiin pureutua oikeuslingvivistisin ja informaatiotutkimuksen keinoin muun muassa tutkimalla tekoälyn mahdollisuuksia luoda uusia oikeudellisia käsitteitä olemassa olevaan käsitejärjestelmään.

Oikeusinformatiikka

Oikeusinformatiikan tutkimuskohteita ovat tiedon ja informaatioteknologian merkitys ja vaikutus oikeudellisissa toiminnoissa, tietojenkäsittelyn ja informaatioteknologian kehittäminen oikeudellisiin toimintoihin sekä oikeudelliset tietoineistot. Oikeusinformatiikan tutkimuksessa erityisesti informaatioteknologian käytön vaikutus lainsäädäntöön sekä sääntelyn ohjaava vaikutus informaatioteknologian käyttöön on merkittävässä asemassa.²³

Informatiikka itsessään viittaa informaation, teknologian ja niiden vaikutuksen tutkimiseen jollakin erityisellä alalla, kuten oikeustai lääketieteessä.²⁴ Oikeusinformatiikan tutkimus painottuuakin informaatioteknologian hyödyntämiseen erilaisissa oikeudellisissa tietoineistoissa ja oikeudellisissa toiminnoissa, ja se rinnastetaan

22 Finto. Suomalainen asiasanasto- ja ontologiapalvelu

23 Magnusson Sjöberg 2021, 30.

24 Hinson 2005, 134.

4.3 Oikeustieteen ja informaatiotutkimuksen yhteyksiä etsimässä

informaatioteknologiaoikeuteen (IT-oikeus),²⁵ vaikka IT-oikeus on oikeudenala tutkimussuuntauksen sijaan.

Oikeusinformatiikan tutkimukselle on tyypillistä, että se saa syötteensä käytännön ongelmista käsin,²⁶ joten sitä ei voida pitää erityisen teoriapainotteisena tutkimusalana. Teorioita hyödynnetään oikeusinformatiikassa, mutta ne saavat syötteensä oikeustieteestä tai oikeusinformatiikkaan liittyvistä muista tieteistä, kuten informaatiotieteestä. Voidaankin todeta, että erityisesti informaatiotutkimus ja tietojenkäsittelytiede näyttävät soveltuvan hyvin tukemaan oikeusinformatiikan tutkimusta.²⁷ Ilman näiden tutkimusalojen yhteyttä oikeudelliseen tutkimukseen ei oikeusinformatiikasta voida puhua omana tutkimusalanana.²⁸

Oikeusinformatiikan tutkimus on moniammatillista²⁹ ja tieteidenvälistä tutkimusta, jossa samaa asiaa tutkitaan usean tieteen keinoin.³⁰ Käsiteltäessä tietojenkäsittelyyn liittyviä kysymyksiä esimerkiksi puhtaasti oikeusdogmaattisesti informaatio-oikeuden tai IT-oikeuden oikeudenaloilla ei kysymys ole oikeusinformatiikan tutkimuksesta, vaan oikeustieteen ytimeen sijoittuvasta oikeusdogmaattisesta tutkimuksesta. Tieteidenvälisyyys tekee oikeudellisesta tutkimuksesta oikeusinformatiikkaan sijoitettavan tutkimuksen silloin, kun oikeudellista informaatiota ja sen käsittelymenetelmiä lähestytään monitieteellisellä näkökulmalla erilaisia tutkimusmenetelmiä yhdistäen.³¹

Informaatiotutkimukseen liittyvää oikeusinformatiikan tutkimusta on tehty muun muassa oikeudellisten asiakirjojen rakenteisuudesta,³² oikeudellisten tietoaaineistojen semanttisuuden ke-

25 Erdelez & O'Hare 1997, 367–370.

26 Kuopus 1988, 23.

27 Hinson 2005, 150.

28 Voutilainen 2009, 14.

29 Saarenpää 2018, 33–34.

30 Karhu 2005, 27.

31 Voutilainen 2009, 14.

32 Dolin 2021b, 61–68.

hittämisestä,³³ oikeudellisen tiedon visualisoinnista ja asiakirjojen suunnittelusta,³⁴ oikeudellisten asiakirjojen automaattisesta laatiemisesta,³⁵ pääsystä oikeudellisiin tietoaaineistoihin,³⁶ oikeudellisen tiedonhaun tehokkaista keinoista,³⁷ oikeudellisen tiedon digitaalisesta käsittelystä,³⁸ tekoälyn hyödyntämisestä oikeudellisten aineistojen käsittelyssä³⁹ sekä tietoturvallisuudesta.⁴⁰

Oikeusinformatiikan tutkimuksen ja sen kehityksen on eräissä avauksissa nähty olevan hiipumassa 2000-luvulla,⁴¹ mutta niin Euroopassa kuin Yhdysvalloissakin, samoin eräiltä osin Aasiassa, julkaistaan varsin paljon tutkimustuloksia oikeusinformatiikan alueella ja nimenomaisesti asemoiden tutkimus oikeusinformatiikkaan. Oikeustieteen informaatiokeskeisyys ja oikeudellisen toiminnan digitalisoituminen nostavat esiin uusia teemoja niin oikeustieteellisen tutkimuksen kuin informaatiotutkimuksen saroilla ja yhteistyössä oikeusinformatiikan tutkimusalustalla.

Oikeudellinen informaatio – muodostuminen ja käyttö

Oikeudellinen informaatio

Yhteiskuntamme oikeusjärjestys muodostuu oikeudellisesta informaatiosta, jota on tuotettu erilaisiin oikeuslähteisiin.⁴² Käytännössä oikeustieteen asiantuntijat, juristit, käsittelevät päivittäin oikeuslähteitä, jotka muodostavat oikeudellisia tietoaaineistoja. Oikeuslähteitä ovat lait, asetukset, viranomaisten antamat määräykset ja säädösten

33 Oksanen ym. 2019, 212–226.

34 Sharma ym. 2021, 221–222.

35 Lauritsen 2021, 69–83.

36 Bannan 2021, 467–480 ja Sharma ym. 2021, 223.

37 Sharma ym. 2021, 224; Alexopoulos ym. 2020, 98–105 ja Virkar ym. 4–15.

38 Malmgren 2021, 541–584.

39 Nay 2023, 309–391.

40 Pöysti 1999, 454–464; Råman 2006 passim.; Andersson & Norden 2021, 67–161.

41 Koulu 2018, 848–852.

42 Ks. oikeuslähteistä ja niiden merkityksestä Määttä 2024a, 24–26.

4.3 Oikeustieteen ja informaatiotutkimuksen yhteyksiä etsimässä

valmisteluaineistot, kuten hallituksen esitykset sekä eduskunnan valiokuntien lausunnot ja mietinnöt. Lisäksi näiden tietoaineistojen soveltamisesta muodostuu oikeuskäytäntöä tuomioistuinten päätöksinä. Erityisesti hallintotoiminnassa merkitystä saavat myös eduskunnan oikeusasiamiehen ja valtioneuvoston oikeuskanslerin antamat laillisuusvalvontaratkaisut sekä muiden laillisuusvalvojien, kuten tietosuojavaltuutetun, antamat ratkaisut. Lisäksi oikeuskirjallisuus tuottaa tietoa lainsäädännön soveltamisen tueksi, ja sen on katsottu kuuluvan sallittuihin oikeuslähteisiin.⁴³ Myös ohjeet ja suositukset ovat sallittuja oikeuslähteitä, mutta ne eivät ole luonteeltaan sitovia, vaikka ohjeen tai suosituksen olisi antanut viranomainen.⁴⁴ Ohjeilla voi kuitenkin olla tosiasiassa myös laintulkintaa ohjaava vaikutuksensa.

Oikeudellisen informaation käsittelyä tehdään erilaisissa oikeudellisissa toimissa. Lainvalmistelussa ministeriöissä selvitetään yhteiskunnan oikeudellista tilaa, siinä olevia mahdollisia puutteita ja tämän tiedonhankinnan perusteella sääntelytarpeita. Lainvalmistelussa kerätään tietoa oikeuskäytännöstä ja erilaisista tutkimuksista sekä teetätetään tutkimuksia ja selvityksiä sääntelyalueen tilasta.⁴⁵ Varsinaisessa säädösten valmistelussa on puolestaan arvioitava erilaisia sääntelyvaihtoehtoja olemassa olevan ja kerätyn tiedon pohjalta. Lainvalmistelu edellyttää paitsi sääntelytekniikan hallintaa myös sääntely-ympäristön ja sen käsitteistön tuntemusta. Lainvalmistelussa on myös tehtävä laadittujen lakiehdotusten vaikutusarviointia, joka on dokumentoitava osaksi lainvalmisteluasiakirjoja.⁴⁶

Lainvalmistelua voi luonnehtia moniammatilliseksi ja -tieteelliseksi prosessiksi, jossa vaaditaan ainakin laajempia lakiehdotuksia valmisteltaessa monipuolista tietojenkäsittelyn ja tiedonhallinnan osaamista. Lainvalmistelun lopputuloksena muodostuu yhtenäinen asiakirja, joka annetaan valtioneuvoston yleisistunnon päätöksellä

43 Aarnio 1989, 221.

44 Määttä 2024b, 2–14.

45 Tala 2005, 87.

46 Niemivuo 2020, 132.

hallituksen esityksenä eduskunnalle. Myös eduskuntakäsittelyssä hallituksen esityksen käsittelemiseksi kerätään monialaista tietoa eri asiantuntijoilta hallituksen esityksen sisällöstä. Lainvalmistelu- ja lainsäädäntöprosessi sisältävät siis runsaasti tiedon hakemista, uuden tiedon muodostamista sekä erilaisten oikeudellisesti merkittävien asiakirjojen tuottamista.

Perustuslain 79 §:n 2 momentissa on säädetty lakien julkaisemisesta säädöskokoelmassa. Säännöksen mukaan valtioneuvoston on lain vahvistamisen jälkeen viipymättä julkaistava laki Suomen säädöskokoelmassa. Tarkemmin säädöskokoelmasta ja siinä tapahtuvasta julkaisusta on säädetty Suomen säädöskokoelmasta annettussa laissa (188/2000), jonka 2 §:n mukaan säädöskokoelmassa julkaistaan lait, eduskunnan työjärjestys ja eduskunnan täysistunnon eduskuntatyön järjestämisestä tai eduskunnan valitsemia toimielimiä varten hyväksymät ohje- ja johtosäännöt sekä tasavallan presidentin, valtioneuvoston ja ministeriöiden asetukset. Säädöskokoelmassa julkaistaan myös valtionaloutta koskevat eduskunnan päätökset. Säädöskokoelmaan kuuluu erillisenä osana sopimussarja, joka sisältää Suomen tekemät valtiosopimukset sekä niihin liittyvät voimaansaattamislait ja -asetukset. Suomen säädöskokoelma onkin keskeinen oikeudellisen tiedon tietovaranto, jossa säädökset ja niiden muutokset on julkaistu alkuperäisinä ja todistusvoimaisina.

Säädöskokoelma on julkaistu sähköisesti. Nykyään säädöskokoelman virallinen versio on sähköinen, ja se on saatavissa yleisessä tietoverkossa oikeusministeriön Finlex-palvelusta. Säädöskokoelmassa julkaistut säädökset ovat todistusvoimaisia alkuperäiskappaleita, joten viime kädessä jonkin säädöksen sisällön sanamuoto onkin tarkastettava säädöskokoelmassa julkaistusta versiosta. Finlex-palvelussa on julkaistu myös tietyn lain ja asetuksen eri muutos-säädöksistä koottu ajantasainen versio, mutta tällaiset koosteet on tuotettu tietopalveluna eivätkä tiedot nauti julkista luotettavuutta. Ne eivät myöskään ole todistusvoimaisia sellaisenaan, vaikka niitä käytetäänkin pääosin oikeuslähteinä säädöskokoelmassa julkaistujen yksittäisten säädösten sijaan.

4.3 Oikeustieteen ja informaatiotutkimuksen yhteyksiä etsimässä

Säädöskokoelma ei ole ainoa säädöksiin tai niihin rinnastettaviin oikeuslähteisiin liittyvä tietovaranto. Viranomaisten antamat määräykset julkaistaan ministeriöiden ja muiden valtion viranomaisten määräyskokoelmista annetun lain (189/2000) 1 §:n mukaan viranomaiskohtaisessa määräyskokoelmassa. Viranomaisilla voi olla yhteinen määräyskokoelma, jos asianomainen ministeriö niin päättää. Euroopan unionin säädökset julkaistaan Euroopan unionin toiminnasta tehdyn sopimuksen 297. artiklan nojalla Euroopan unionin virallisessa lehdessä. Euroopan unionin virallisen lehden sähköisestä julkaisemisesta annetun neuvoston asetuksen (EU N:o 216/2013) mukaan vain sähköisessä muodossa julkaistu virallinen lehti on todistusvoimainen ja vain sillä on oikeusvaikutuksia. Virallisen lehden sähköinen versio asetetaan yleisesti saataville EUR-Lex-portaaliin vanhentumattomassa muodossa ja rajoittamattoman ajan. EUR-Lex-portaalissa julkaistaan myös eri säädösten konsolidoituja toisintoja, jotka sisältävät ajantasaisen koosteen tietyn säädöksen, kuten Euroopan unionin asetuksen tai direktiivin sisällöstä, mutta ne eivät ole todistusvoimaisia.⁴⁷

Juristien ammattikunnassa on vahvasti tukeuduttu niin kutsuttuun lakikirjaan, joita julkaisee Suomessa muutama kustantaja. Lakikirjat eivät ole kuitenkaan virallisia asiakirjoja, vaan toimitettuja teoksia, joissa säädökset on koostettu kustantajan haluamalla tavalla esimerkiksi jonkin tietyn teeman tai oikeudenalan ympärille. Lisäksi tällaiset teokset sisältävät tyypillisesti viittauksia pykäläkohtaiseen oikeuskäytäntöön. Näitä lakikirjoja voidaankin kutsua jonkinlaisiksi hakuteoksiksi. Sääntelymäärän kasvaessa ja lainsäädännön muutosten nopeutuessa lakikirjat voivat olla jo julkaisuhetkellä osittain vanhentuneita.

Oikeuslähteiden tietolähteet – informaatiolähteet – ovat kelvollisia juristin työssä silloin, kun ne ovat saatavilla ja niistä on haettavissa tietoa helposti oikeudellisten ongelmien ratkaisemisen tueksi luotettavasti.

⁴⁷ Miettinen 2013, 415.

Oikeudellisen informaation saatavuus

Oikeusvaltion ja perusoikeuksien toteuttamisen kannalta katsottuna oikeudellisen informaation saatavuuden vaatimuksen voidaan katsoa olevan suoraan johdettavissa perustuslain 12 §:n 2 momentissa säädetystä asiakirjajulkisuudesta sekä 21 §:ssä säädettyistä hyvän hallinnon ja oikeusturvan vaatimuksista. Lähtökohtainen oikeus saada tietoa oikeudenkäyntiasiakirjoista turvaa osaltaan tuomiovallan käytön julkisen arvioinnin sekä lisää luotettavuutta valan käyttöön.⁴⁸ Oikeudellisen informaation saatavuudella on sekä henkilö- että teknologiaulotteinen näkökulma. Oikeudellinen informaatio tulee olla jokaisen saatavilla saavutettavassa muodossa, jotta jokainen voi saada siihen perehtymällä selon omista eduistaan, oikeuksistaan ja velvollisuuksistaan. Oikeudellisen informaation saantiin liittyvät puutteet voivat estää yksilön oikeusturvan toteutumisen.⁴⁹ Oikeudellisen informaation tulee olla myös helposti saatavilla, jolloin erityinen merkitys on oikeudellisen tiedon esitystavalla ja löydettävyydellä tietoverkosta esimerkiksi hakukoneiden avulla.⁵⁰

Kuitenkaan kaikki oikeudellinen informaatio ei ole helposti saatavilla. Tuomioistuimet ovat monopolisoineet tuomioiden julkistamisen, ja erityisesti käräjäoikeuksien tuomioista ei ole saatavissa kuin yksittäisiä tietoja oikeuskäytännöstä. Toisaalta oikeuskäytännössä on todettu, että joukkotiedotusvälineillä on olennainen rooli oikeudenkäynnin julkisuuden tosiasiallisessa toteutumisessa,⁵¹ vaikka viranomaisten toiminnan julkisuudesta annetun lain (621/1999, julkisuuslaki) 20 §:ssä on säädetty viranomaisen, kuten tuomioistuimen, tehtäväksi tuottaa ja jakaa tietoa ratkaisukäytännöstä.

Tuomioistuimissa muodostuu paljon merkityksellisiä tietoa-aineistoja, jotka kuvaavat yhteiskunnan oloja niin rikollisuudesta kuin myös julkisen vallan käytöstä ja yksityisten välisistä suhteista.

48 Korkeimman oikeuden ennakkopäätös KKO 2023:69, kohta 8.

49 Dolin 2021a, 15.

50 Dolin 2021a, 14–15.

51 KKO 2023:69, kohta 10.

Haasteena tietojen saatavuudessa ovat henkilötietojen suojaa koskevat kysymykset sekä resurssit, joilla oikeudenkäyntiaineistoja tuotetaan julkaistavaan muotoon. Joukkotiedotusvälineiden varaan rakentuva tiedon tuottaminen tuomiovallan käytöstä on valikoivaa otsikkojournalismia. Oikeudenkäyntiasiakirjojen saatavuus ei ole pelkästään kansallinen ongelma, vaan yleinen oikeudellisen informaation hallintaan liittyvä tutkimustyötä ja informaatioteknologian kehittämistäkin edellyttävä kohde.⁵²

Oikeudellinen sääntely ja asiakirjat

Lainsäädännön nojalla laaditaan erilaisia yksityis- ja julkisoikeudellisia asiakirjoja. Tätä säädettyjen asiakirjojen laatimisvelvollisuutta voidaan kutsua yleisesti dokumentointivelvollisuudeksi, joka koskee oikeustoimien tekemistä tai hallintotoimien suorittamista sekä julkisen vallan käyttöä. Dokumentointivelvollisuudella on merkitystä erilaisten toimien todennettavuuden kannalta. Yksityisoikeudellisten asiakirjojen laatimisesta on säädetty muun muassa perhe- ja perintöoikeuden sekä sopimusoikeuden sääntelyssä. Julkisoikeudellisessa sääntelyssä on säädetty muun muassa erilaisten päätösten sisällöstä sekä yleisemmin asiakirjojen hallinnan ja tietoturvallisuuden vaatimuksista.

Yhteiskuntamme perusrakenteet ja -yksiköt on tallennettu perustietovarantoihin,⁵³ joihin tehdyt merkinnät muodostavat erilaisia oikeuksia niin yksilöille kuin yhteisöillekin. Viranomaisissa asiakirjoista muodostuvat tietoaaineistot luovat merkinnöillään oikeudellisesti julkisorganisaatioille, yksilöille ja yhteisöille erilaisia etuja, oikeuksia tai velvollisuuksia. Valtion hallintoviranomaiset tekevät hallintopäätöksiä, laativat muistioita, tekevät määrärahojen käyttöä koskevia ehdotuksia ja laativat erilaisia suunnitelmia. Kunnan viranomaiset tekevät toimielimissä päätöksiä, jotka dokumentoidaan pöytäkirjoihin. Tuomioistuimet puolestaan ratkaisevat sinne vireille

52 Magrath 2015, 190–193.

53 Kuronen 1998, 52.

tulleita lainkäyttöasioita, joita annetaan ratkaisuihin ja tuomioina. Kaiken kaikkiaan lainsäädännön täytäntöönpano tuottaa erittäin runsaasti oikeudellista informaatiota, joka on myös merkityksellistä informaatiotutkimukselle.

Oikeudelliseen informaatioon liittyvät haasteet

Yhteiskunnassamme informaatiotutkimuksen kannalta on oma haasteensa siinä, ettei lainsäädännössä ole säädetty siitä, miten esimerkiksi viranomaisten sähköisten asiakirjojen alkuperäisyys ja siten todistusvoimaisuus varmistetaan. Sähköisestä asiakirjasta voi olla useita kappaleita, mutta asiakirjan alkuperäiskappaleesta ei ole välttämättä minkäänlaista merkintää tai tietoa. Esimerkiksi lausuntomenettelyssä saadut sähköisessä muodossa olevat lausunnot saatetaan tallentaa useampaan tietojärjestelmään eri tarkoituksia varten. Sähköisesti annettu lausunto alkuperäisenä asiakirjana on kuitenkin siinä tietojärjestelmässä, johon se on lähetetty tai tallennettu lausunnon antamista varten, jos se on laadittu alkuperäisesti sähköisessä muodossa. Alkuperäinen kappale voi olla myös lausunnon laatineella viranomaisella, jolloin lausunnosta on toimitettu kopio lausuntoa pyytäneelle viranomaiselle. Puolestaan paperilta skannatut asiakirjat ovat jäljennöksiä eikä niiden alkuperäisyyttä tyypillisesti voida varmistaa muuta kuin paperiversiosta, jota ei välttämättä edes säilytetä.

Asiakirjojen tuottamiseen ja hallintaan liittyy niin hyvän hallinnon kuin oikeusturvankin kannalta merkityksellisiä vaatimuksia julkisessa hallinnossa. Esimerkiksi perustuslain 118 §:n mukaan virkamies vastaa virkatoimiensa lainmukaisuudesta. Tästä syystä virkatoimien toteuttamista osoittavista asiakirjoista on tultava ilmi, kuka virkatoimen on suorittanut, jotta voidaan todentaa, onko virkatoimen suorittanut toimivaltainen henkilö ja kuka virkatoimesta on vastuussa.⁵⁴ Perustuslakivaliokunta on pitänyt hyvän hallinnon kannalta tarpeellisenä, että viranomaisten päätöksistä ilmenee niistä

54 Voutilainen & Ouli 2023, 62.

vastuussa olevien henkilöiden nimet.⁵⁵ Asiakirjojen hallintaan ja erilaisten toimien todennettavuuteen liittyvät myös lainsäädännössä olevat vaatimukset asiakirjojen säilyttämisestä. Esimerkiksi julkisen hallinnon tiedonhallinnasta annetun lain (906/2019, jäljempänä tiedonhallintalaki) 21 §:ssä on säädetty yleisistä kriteereistä, joiden perusteella viranomaisille muodostuu tietoaaineistojen säilytysvelvollisuus. Puolestaan arkistoinnista säädetään arkistolaissa (831/1994). Arkistoinnin keskeisenä tarkoituksena on yhteiskunnan toiminnassa muodostuneen keskeisen informaatioperinnön säilyminen erityisesti tutkimustarkoituksia varten. Oikeudellinen sääntely muodostaa pääasiallisen pohjan yhteiskunnan asiakirjallisen tietopääoman muodostumiselle ja siten muun muassa informaatiotutkimuksen edellytyksille.

Oikeudellisen informaation kasvu tulee jatkumaan, koska yhteiskunnan oikeudellistuminen eli yhteiskunnan toimintojen ohjaaminen sääntelykeinoin on tullut jäädäkseen.⁵⁶ Suomalaisessa oikeusjärestyksessä erityislainsäädäntöön tukeutuminen aiheuttaa oikeudellisen informaation hallinnalle haasteita käytännön soveltamistilanteissa.⁵⁷ Yleis- ja erityislakien väliset suhteet eivät ole välttämättä selviä, ja tietyn oikeudellisen ongelman ratkaisemiseen tarvitaan erityisesti julkisoikeudessa useaan säädökseen sisältyvien säännösten samanaikaista tulkintaa ja hallintaa. Oikeudellisen informaation hallintaan ei ole kuitenkaan toistaiseksi esitetty tutkimuksellisesti kestäviä ratkaisuja sääntelyverkoston informaation hallinnoinnin kehittämiseksi.

Oikeudelliset tietoaaineistot ja -varannot muodostavat merkittävän yhteiskuntamme elinoloja ja toimintaa kuvaavan kokonaisuuden niin nykyajan ilmiöiden tutkimiseen mutta myös historian tutkimukselle. Yhteiskunnan tietyn ajan ja yhteiskuntajärjestyksen ymmärtäminen edellyttää myös kunkin ajanhetken sääntely-ympäristöön paneutumista. Tästä syystä oikeudelliset tietoaaineistot ja

⁵⁵ PeVL 38/2021 vp, 4.

⁵⁶ Koivumaa 1995, 545–546.

⁵⁷ Pöysti 2006, 327.

-varannot on muodostuttava luotettavasti ja niiden tulee olla tutkijoiden käytettävissä.

Informaatio-oikeus – sääntelykehikko informaation käsittelylle

Informaatio-oikeus on oikeudenala, jonka sääntelyn kohteena ovat tieto sekä sen käsittely ja hallinta. Oikeudenalana informaatio-oikeus muodostaa sääntely-ympäristön tiedolle ja siihen liittyville toiminnoille eli tietojenkäsittelylle ja tiedonhallinnalle. Informaatio-oikeuden ytimessä on tietoon liittyvien oikeuksien sääntelyjärjestelmä. Informaatio-oikeuden sääntelyssä määritellään erityisesti yksilöiden ja yhteisöjen oikeutta tietoon, ja sääntely vastaa kysymyksiin siitä, kenellä on oikeus tietoon, mihin tietoon ja missä tilanteissa.⁵⁸ Informaatio-oikeudellisessa sääntelyssä tasapainoillaan tiedollisen yksityisyyden, yhteiskunnan tietotarpeiden, tiedollisten omistus- ja hallintaoikeuksien ja tiedonsaantiin liittyvien oikeuksien sääntelyrakenteissa, joissa säädetään tietoon liittyvistä oikeuksista ja niiden käytön kontrollista.⁵⁹

Informaatio-oikeuden sääntelykohteena on tieto eri muodoissaan. Tietoa voidaan pitää sisäistettynä informaationa. Säädoskoelma pitää sisällään informaatiota. Yksittäinen laki tai säännös on informaatiota, mutta lain ja sen säännösten sisäistäminen ja ymmärtäminen muuttaa informaation tiedoksi tulkitsijalleen.⁶⁰

Tiedon arvonmääritys on käsitteenä monimerkityksellinen,⁶¹ mutta sitä voidaan pitää prosessina, jossa tietoaaineistosta tunnustetaan asiakirjoille säilytysarvo erityisesti arkistotarkoituksia varten. Tiedon arvon määrittäminen ja yleensäkin tiedon arvo on vaikeasti hahmotettava monisyinen kokonaisuus, joka on ollut pitkään myös

58 Voutilainen 2019, 24.

59 Lipton 2003, 707–778 ja Voutilainen 2019, 21.

60 Pihlajamäki 2004, 22–36 ja Bing 1986, 52–55.

61 Valtonen 2015, 95–98.

informaatiotutkimuksen kohteena monitieteellisessä tarkastelussa.⁶² Kuitenkin oikeudellisen sääntelyn taustalla ovat myös erilaiset tiedon arvoon vaikuttavat tekijät, joita halutaan sääntelyllä ohjata ja suojata. Tiedolla voi olla rahallinen arvo, jonka suojaksi on säädetty muun muassa tekijänoikeuksista. Tietoa voidaan antaa tai tuottaa vastiketta vastaan. Tiedolle voi muodostua ajan saatossa historiallinen arvo, jota säilytetään arkistoinnin avulla, minkä vuoksi arkistointivelvollisuudesta on säädetty lailla. Tiedolla on myös varmuus- ja todistusarvo, jolla voidaan osoittaa yksilöiden ja yhteisöjen erilaisia etuja, velvoitteita ja oikeuksia. Tästä syystä tietoon kohdistuu säilytysvelvollisuuksia sekä velvollisuuksia suojata tietoa tietoturvallisuustoimenpiteillä tiedon eheyden, luottamuksellisuuden ja käytettävyyden varmistamiseksi. Tieto on vallankäytön väline ja osa vallankäyttöä. Mitä harvemman käytettävissä tieto on, sen arvokkaampaa se on sekä rahalliselta arvoltaan että vallankäytön välineenä. Tätä vasten julkisen vallan käyttöä ja sen avoimuutta ja valvontaa varten on säädetty julkisuusperiaatteesta ja sen toteuttamisesta estämään julkisen vallan ja julkisten varojen väärinkäyttöä.⁶³

Oikeustieteissä oikeusjärjestys on jaettu tieteenalalla muodostuneiden käytäntöjen ja lainsäädännön kehittymisen kautta oikeudenaloihin. Oikeustieteessä oikeudenalajaotus palvelee osittain tiedonhallinnallisia tarkoituksia siten, että sääntelymassa on jäsennetty oikeudenaloihin, jotka muodostuvat oikeudenalan säännöksistä, käsitteistä ja periaatteista sekä sääntelyä koskevasta soveltamiskäytännöstä. Kukin oikeudenala saa kiinnekohtansa tietystä sääntelykohteesta, kuten informaatio-oikeudessa – tiedosta, tietojenkäsittelystä ja tiedonhallinnasta.

Oikeudenaloihin kiinnittyvä tiedonhallinnallinen jäsenitys näkyy perinteisissä lakikirjoissa, jotka on sisällöllisesti jaettu perinteisesti pääosin oikeudenaloittain. Oikeudenalajaotuksella on erilaisia tehtäviä. Sillä voidaan paikantaa oikeudellisia ongelmia oikeusjär-

⁶² Repo 1984, 51–59.

⁶³ Voutilainen 2019, 22.

jestyksestä (heuristinen tehtävä). Oikeudenalajaotuksen tehtävänä on jäsentää lakimieskoulutusta hallittaviin kokonaisuuksiin (pedagoginen tehtävä).⁶⁴ Lisäksi oikeudenalajaotuksella jäsennetään säädösvalmistelun vastuita esimerkiksi valtioneuvostossa sekä eduskunnassa (organisatorinen tehtävä). Oikeudenalajaotus ei ole kuitenkaan niin kutsutusti kiveen hakattu, vaan se elää yhteiskunnan muuttuessa ja oikeusjärjestyksen kehittyessä.

Informaatio-oikeus oikeudenalanakin on varsin tuore osa oikeusjärjestyksen systematisointia, mutta sen varhaiset juuret sijoittuvat jo 1970-luvulle.⁶⁵ Varsinaisesti informaatio-oikeuden systematisointia ryhdyttiin tekemään 1980–1990-lukujen taitteessa.⁶⁶ Syynä informaatio-oikeudellisen sääntelyn ja oikeudenalan kehittymiselle ovat olleet yhteiskunnalliset muutokset. Informaatiosta muodostetun tiedon perusteella suunnitellaan ja ohjataan yhteiskunnan eri toimintoja. Tässä yhteydessä puhutaan tyypillisesti tietojohdamisesta tai tiedolla johtamisesta. Myös informaatioteknologian kehityksellä on ollut roolinsa informaatio-oikeuden sääntelyn kehitykselle erityisesti yksityisyyden suojaa koskevan sääntelyn kehittymisen kannalta,⁶⁷ vaikka IT-oikeus muodostaa oman oikeudenalansa, jonka sääntelykohteena ovat informaatioteknologia ja tietoverkot sekä niiden käyttöedellytykset yhteiskunnassa.⁶⁸

Koska tietoa voidaan hyödyntää erilaisissa yhteiskunnan toiminnoissa ja tiedot kohdistuvat monin osin yksilöihin, on yksilön suojaaminen henkilötietojen käsittelyssä noussut tärkeään asemaan. Niin ikään erilaisesta informaatiosta muodostettu taloudellista arvoa sisältävä tietoaaineisto tai tieto tarvitsee suojaa, minkä vuoksi tekijänoikeuksilla on merkitystä hallinnoitaessa teoskynnyksen ylittävää informaatiota. Yhteiskuntamme toiminta perustuu sen informaation varaan, jota eri tietovarantoihin on tallennettu. Tämän

64 Tuori 2002, 107–109.

65 Gasser 2002, 8.

66 Dommering 1991, 27–28.

67 Trubow 1985, 815–816.

68 Voutilainen 2009, 344.

tiedon hallinta edellyttää niin kutsutusti pelisääntöjä, jotka on vahvistettava lainsäädännössä. Näitäkin tietoja on suojattava yhteiskunnan toiminnan varmistamiseksi, ja tietoturvallisuussääntely on yksi keino tämän suojan toteuttamiseksi.

Euroopan unionin sääntely vaikuttaa keskeisesti informaatio-oikeudelliseen sääntelyyn, koska tieto on markkinahyödyke Euroopan unionin sisämarkkinoilla. Tiedon liikkuvuus ja sen turvaaminen Euroopan unionin jäsenvaltioiden välillä liittyy erityisesti ihmisten ja palvelujen liikkuvuuden vapauden turvaamiseen. Esimerkiksi Euroopan unionin yleisen tietosuoja-asetuksen tavoitteena on henkilötietojen vapaan liikkuvuuden turvaaminen Euroopan unionin alueella. Tietosuoja-asetuksen 1 (3) artiklan mukaan henkilötietojen vapaata liikkuvuutta unionin sisällä ei saa rajoittaa eikä kieltää syistä, jotka liittyvät luonnollisten henkilöiden suojeluun henkilötietojen käsittelyssä. Esimerkiksi tietosuojaan liittyvillä syillä ei voida kieltää unionin jäsenvaltioiden rajat ylittävää informaatiotutkimusta, jossa käsitellään henkilötietoja.

Tutkijan oikeus tietoon

Tutkija henkilötietoja sisältävän informaation käsittelijänä

Informaatio-oikeudellinen sääntely vaikuttaa keskeisesti informaatiotutkijan tutkimuksen tekemiseen ja tutkimuksen tekemisen edellytyksiin. Informaatio-oikeudellisen sääntelyn tuntemusta voidaan pitää informaatiotutkijan ja -ammattilaisen pätevyyden perusvaatimuksena. Perustuslain 16 §:n 3 momentissa tieteen vapaus on turvattu jokaisen perusoikeutena. Tieteen vapauten sisältyy tieteellisen tutkimuksen vapaus. Tieteen vapauden keskeisenä sisältönä on luoda tieteenharjoittajalle vapauspiiri, jonka puitteissa hänellä on oikeus päättää tieteellisen tiedon hankintaa, tulkintaa, arviointia, leviättämistä ja omaksumista koskevista ratkaisuista.⁶⁹ Perustuslain 16 §:n 3 momentin sääntelystä johtuen lainsäätäjän on lainsäädännössä

69 Miettinen 2001, 277.

turvattava edellytykset, joilla tieteellisen tutkimuksen vapautta voidaan toteuttaa.

Tieteelliselle tutkimukselle ei ole kuitenkaan laissa säädetty erityistä tunnusmerkistöä, vaikka tieteellinen tutkimus on myös informaatio-oikeudellisessa sääntelyssä sääntelykohteena. Tieteellisen tutkimuksen tunnusmerkistöä onkin luotu oikeuskäytännön ja -kirjallisuuden perusteella. Tutkimuksen tieteellisyyttä arvioidaan oikeudellisesti asianmukaisen tutkimussuunnitelman, tutkimuksen tiedeyhteisöön sidonnaisuuden, tutkimustradition, tutkimustulosten arvioitavuuden, tutkimuksen autonomisuuden ja julkisuuden kannalta katsottuna.⁷⁰

Informaatiota käsittelevän tutkijan tulee tuntea sitä tutkimustietoa-aineistoa koskeva sääntely, jota hän aikoo käsitellä tutkimuksessaan, koska sääntely paitsi mahdollistaa myös rajoittaa tutkimuksen kohteena olevien aineistojen saantia ja käsittelyä. Kun tutkija käsittelee henkilötietoja sisältävää informaatiota, hänen on huomioitava henkilötietojen suojaa koskeva sääntely, joka asettaa myös tieteellisessä tutkimuksessa tapahtuvalle tietoa-aineistojen käsittelylle vaatimuksia. Euroopan unionin yleinen tietosuojaa-asetus tulee sovellettavaksi tehtäessä tieteellistä tutkimusta. Asetuksen 89 (1) artiklan mukaan yleisen edun mukaisia tieteellisiä ja historiallisia tutkimustarkoituksia varten tapahtuvaan käsittelyyn sovelletaan rekisteröidyn oikeuksia ja vapauksia koskevia asianmukaisia suojatoimia tietosuojaa-asetuksen mukaisesti. Näillä suojatoimilla on varmistettava, että rekisterinpitäjä on toteuttanut tekniset ja organisatoriset toimenpiteet, joilla taataan etenkin tietojen minimoinnin periaatteen noudattaminen. Asetuksen johdanto-osan 159 kohdan mukaan tieteellisiä tutkimustarkoituksia varten tapahtuvaa henkilötietojen käsittelyä on tulkittava laajasti niin, että se tarkoittaa myös teknologian kehittämistä ja esittelyä, perustutkimusta, soveltavaa tutkimusta ja yksityisin varoin rahoitettua tutkimusta. Varsinaisesti tutkimustarkoituksissa tapahtuva henkilötietojen käsittely sijoittuu

70 Voutilainen 2015, 5–7.

yleisen edun mukaiseen käsittelyperusteeseen, josta on säädetty kansallisen tietosuojalain (1050/2018) 4 §:n 3 kohdassa.

Tietosuojasääntelyssä on useita säännöksiä, joiden tarkoituksena on mahdollistaa henkilötietojen käsittely yleisen edun mukaisessa tieteellisessä tutkimuksessa. Tietosuoja-asetuksen 5 (1) artiklan b alakohdan mukaan myöhempää henkilötietojen käsittelyä yleisen edun mukaisia tieteellisiä tai historiallisia tutkimustarkoituksia varten ei katsota yhteensopimattomaksi alkuperäisen henkilötietojen käsittelyn tarkoitusten kanssa. Säännös tarkoittaa, että käsiteltäessä henkilötietoja tieteellistä tutkimusta varten ei käsittely ole käyttötarkoitussidonnaisuuden vaatimuksen vastaista, vaikka henkilötiedot olisi kerätty ensisijaisesti muuta kuin tutkimuskäyttöä varten. Tietosuojasääntely ei siten aseta yleistä estettä henkilötietojen keräämiselle erilaisista jo aiemmin muodostetuista tietoaaineistoista, kun käsittelyn oikeusperusteena on tieteellinen tutkimus. Myös erityisiin henkilötietoryhmiin kuuluvia tietoja voidaan käsitellä tieteellisessä tutkimuksessa tietosuoja-asetuksen 9 (2) artiklan j alakohdan ja tietosuojalain 6 §:n 1 momentin 7 kohdan perusteella. Eräissä tilanteissa myös rekisteröidyn oikeuksia voidaan rajoittaa tieteellistä tutkimusta varten tapahtuvassa henkilötietojen käsittelyssä. Tällöin tutkimuksen on kuitenkin täytettävä tietosuojalain 31 §:n 1 momentissa säädetty edellytykset, joiden mukaan rekisteröidyn oikeuksista voidaan poiketa, jos

1. käsittely perustuu asianmukaiseen tutkimussuunnitelmaan
2. tutkimuksella on vastuuhenkilö tai siitä vastaava ryhmä
3. henkilötietoja käytetään ja luovutetaan vain historiallista tai tieteellistä tutkimusta taikka muuta yhteensopivaa tarkoitusta varten sekä muutoinkin toimitaan niin, että tiettyä henkilöä koskevat tiedot eivät paljastu ulkopuolisille.

Näiden säädettyjen ehtojen on täyttyävä samanaikaisesti. Lisäksi jos tutkimuksessa käsitellään erityisiin henkilötietoryhmiin kuuluvia tietoja, on ennen tutkimuksen tekemistä tehtävä tietosuoja-ase-

tuksen 35 artiklassa tarkoitettu tietosuojan vaikutustenarviointi. Tämä arviointi on toimitettava tietosuojalain 31 §:n 3 momentin nojalla tiedoksi tietosuojavaltuutetulle ennen käsittelyyn ryhtymistä.

Tutkijan asema henkilötietojen käsittelyssä

Tutkimuksen toteuttamistavasta riippuen tieteellisessä tutkimuksessa henkilötietojen käsittelystä voi vastata ainakin tutkija, tutkimusryhmä, tutkimuslaitos, yliopisto tai ammattikorkeakoulu. Tietosuoja-asetuksen 24 (1) artiklan mukaan rekisterinpitäjä on kokonaisvastuussa henkilötietojen käsittelystä siten, että rekisterinpitäjän on toteutettava tarvittavat tekniset ja organisatoriset toimenpiteet, joilla voidaan varmistaa ja osoittaa, että käsittelyssä noudatetaan tietosuoja-asetusta. Ennen tutkimuksen käynnistämistä on merkityksellistä määritellä, kuka toimii tutkimuksessa muodostuvan henkilötietoja sisältävän tietoaineiston rekisterinpitäjänä.

Koska tutkimuksen rekisterinpitäjyydestä ei ole säädetty kansallisessa lainsäädännössä, määräytyy rekisterinpitäjyys suoraan tietosuoja-asetuksessa säädetyn rekisterinpitäjyyttä osoittavan määritelmän mukaisesti. Tietosuoja-asetuksen 4 artiklan määritelmäluettelon 7 kohdan mukaan rekisterinpitäjällä tarkoitetaan luonnollista henkilöä, oikeushenkilöä, viranomaista, virastoa tai muuta elintä, joka yksin tai yhdessä toisten kanssa määrittelee henkilötietojen käsittelyn tarkoitukset ja keinot. Näin ollen merkitystä on sillä, kuka viime kädessä päättää tutkimuksen tarkoituksesta ja tutkimuksessa käsiteltävien henkilötietojen keinoista eli tutkimusmenetelmistä. Palvelussuhteella tai taloudellisilla mahdollisuuksilla täyttää rekisterinpitäjälle kuuluvat velvollisuudet ei ole merkitystä määriteltäessä rekisterinpitäjää, vaan määrittely tehdään nimenomaisesti tietosuoja-asetuksessa säädettyä vasten. Tutkimustyön organisoitumisella on siten hyvin paljon merkitystä henkilötietojen käsittelyyn liittyvien vastuiden määrittelyssä. Perustuslaissa turvattuun tieteen vapauden perusoikeuteen kuuluu, että tieteenharjoittajalla on oikeus

valita tutkimusaiheensa ja -menetelmänsä.⁷¹ Informaatiotutkijan onkin syytä perehtyä tietosuojaa koskevaan sääntelyyn kattavasti, jos tutkija on rekisterinpitäjän asemassa. Rekisterinpitäjyys tuo useita velvollisuuksia ja vastuita henkilötietojen käsittelyssä, kuten käsitteilyn osoitus- ja dokumentointivelvollisuudet.

Julkisuuslaki tieteellisen tutkimuksen vapauden turvaajana

Julkisuuslaissa on säädetty julkisuusperiaatteen toteuttamisesta viranomaistoiminnassa. Julkisuuslain keskeisenä tarkoituksena on säätää tiedonsaantioikeuksista ja niiden toteuttamisesta viranomaisen asiakirjoihin. Perustuslain 12 §:n 2 momentin mukaan viranomaisen hallussa olevat asiakirjat ja tallenteet ovat julkisia, ellei niiden julkisuutta ole laissa erikseen rajoitettu. Asiakirjajulkisuus kohdistuu myös viranomaisen tietojärjestelmissä oleviin tietoihin,⁷² joista informaatiotutkimuksen yhteydessä monesti puhutaan rekisteritietoina.

Julkisuuslain 9 §:n 1 momentissa on säädetty jokaisen oikeudesta saada tieto viranomaisen julkisesta asiakirjasta. Julkisuuslain 13 §:n 1 momentin mukaisesti tutkijan esittäessä pyynnön saada tieto viranomaisen julkisesta asiakirjasta ei tutkijalla ole lähtökohteisesti velvollisuutta selvittää tietojen käyttötarkoitusta. Jos pyyntö koskee salassa pidettävää asiakirjaa, tietoa viranomaisen henkilörekisteristä tai muuta asiakirjaa, josta voidaan antaa tieto vain tietyin edellytyksin, tutkijan on julkisuuslain 13 §:n 2 momentin nojalla ilmoitettava tietojen käyttötarkoitus sekä muut tietojen luovuttamisen edellytysten selvittämiseksi tarpeelliset seikat ja tarvittaessa tiedot siitä, miten tietojen suojaus on tarkoitus järjestää.

Tutkijalla on oikeus saada tieteelliseen tutkimukseen viranomaisen henkilörekisterissä olevia julkisia tietoja julkisuuslain 16 §:n 3 momentissa tarkoitettulla tavalla kopiona, tulosteena tai sähköisessä muodossa, jos tiedot ovat tarpeellisia tutkimuksen tekemiseksi. Jos

⁷¹ PeVM 25/1994 vp, 36.

⁷² PeVL 3/2009 vp, 2.

tutkija haluaa tutkia viranomaisen luona henkilötietoja sisältäviä julkisia asiakirjoja tai tallenteita, sovelletaan tietojen antamiseen julkisuuslain 16 §:n 1 momentissa säädettyä. Näissä tilanteissa viranomainen ei voi edellyttää tutkijaa hakemaan tutkimuslupaa, koska tiedonsaanti perustuu yleisöjulkisuutta tarkoittavaan sääntelyyn eikä siinä ole säädetty tutkimuslupamenettelystä. Jos viranomainen vaatii tutkimuslupaa esimerkiksi sillä perusteella, että pyydetyt tiedot ovat henkilötietoja, menettelee viranomainen julkisuuslain vastaisesti toimivaltansa ylittäen.⁷³

Tieteellisen tutkimuksen vapaus on turvattu julkisuuslaissa myös salassa pidettäviin viranomaisen asiakirjoihin tietyissä tapauksissa. Julkisuuslain 28 §:n 1 momentin mukaan viranomainen voi, jollei lailla toisin säädetä, antaa yksittäistapauksessa luvan tietojen saamiseen salassa pidettävästä asiakirjastaan tieteellistä tutkimusta varten, jos on ilmeistä, ettei tiedon antaminen loukkaa niitä etuja, joiden suojaksi salassapitovelvollisuus on säädetty. Lainkohdan mukaan lupaa harkittaessa on huolehdittava siitä, että tieteellisen tutkimuksen vapaus turvataan. Julkisuuslain sääntelyn perusteella viranomainen voi edellyttää tutkijalta tutkimusluvan hakemista ainoastaan tilanteissa, joissa tutkija pyytää saada nähtäväksi salassa pidettäviä asiakirjoja tai pyytää saada pääsyn tutkimaan asiakirjoja johonkin tietoaaineiston säilytystilaan, jossa säilytetään myös salassa pidettäviä asiakirjoja. Jos tutkija pyytää asiakirjoista vain julkisen osan, ei tutkimuslupamenettely tule sovellettavaksi, koska kysymys ei ole pyynnöstä, joka koskisi salassa pidettävien tietojen saantia. Julkisuuslain 28 §:n 2 momentin mukaan lupa voidaan antaa määräajaksi, ja siihen on liitettävä yleisen ja yksityisen edun suojaamiseksi tarpeelliset määräykset. Lupa voidaan tarvittaessa peruuttaa, jos siihen harkitaan olevan syytä.

Tietojen saannista arkistosta on säädetty eräiltä osin erikseen, jos pyyntö saada tieto koskee salassa pidettävää asiakirjaa. Julkisuuslain 27 §:n mukaan arkistolaissa säädettyssä järjestyksessä arkistoon

73 AOA EOAK/101/2019, 8.6.2020.

siirretystä, salassa pidettäväksi säädetystä viranomaisen asiakirjasta saa antaa tietoja tutkimusta varten, jollei asiakirjan siirtänyt viranomainen ole toisin määrännyt. Tietojen antamista harkittaessa on huolehdittava siitä, että tieteellisen tutkimuksen vapaus turvataan. Sääntely tulee sovellettavaksi tilanteissa, joissa asiakirja on tosiasiassa siirretty arkistoviranomaisen, kuten Kansallisarkiston, haltuun.⁷⁴ Julkisuuslain 27 §:n 2 momentin mukaan arkistosta asiakirjat saaneen on annettava kirjallinen sitoumus siitä, ettei hän käytä asiakirjaa sen henkilön vahingoksi tai halventamiseksi, jota asiakirja koskee, tai hänen läheisensä vahingoksi tai halventamiseksi taikka sellaisten muiden etujen loukkaamiseksi, joiden suojaksi salassapitovelvollisuus on säädetty. Mainittua 2 momentin säännöstä ei voida soveltaa siten, että kriittinen tieteellinen tutkimus estyy esimerkiksi tutkittaessa julkisen vallan käyttäjien toimintaa arkistossa olevien asiakirjojen perusteella.

On myös syytä huomata, ettei viranomaiselle lähetettyyn kyselytutkimukseen sovelleta julkisuuslain säännöksiä, ellei kyselytutkimuksessa tehdä yksiselitteisesti pyyntöä saada tietty viranomaisen asiakirja. Siten viranomaisella ei ole toimivaltaa vaatia tutkijaa tekemään tutkimuslupahakemusta, jos tutkija lähettää viranomaiselle kyselytutkimukseen osallistumista koskevan pyynnön. Lailisuusvalvonnassa omaksutun käytännön mukaan perustuslain 21 §:ssä turvatun hyvän hallinnon takeisiin kuuluu, että jokaisella on oikeus saada ilman aiheetonta viivytystä vastaus viranomaiselle lähetettyihin asiallisiin kirjeisiin, tiedusteluihin ja pyyntöihin, joihin kirjoittaja selvästi odottaa vastausta.⁷⁵ Tämä ei kuitenkaan tarkoita sitä, että viranomaisella olisi yleinen velvollisuus vastata kaikkiin sille osoitettuihin kirjoituksiin. Viranomaisella ei ole velvollisuutta vastata varsinaisesti sille tulleeisiin kyselytutkimuksiin, vaan niihin vastaaminen on viranomaisen harkinnassa ottaen huomioon muun muassa viranomaisen käytössä olevat henkilöresurssit.

⁷⁴ Voutilainen 2021, 17–18.

⁷⁵ AOA 4461/4/14 ja AOK OKV/471/10/2021.

Tieteellistä tutkimusta varten kerättävien tietojen saannista on säädetty myös esimerkiksi sosiaali- ja terveystietojen toissijaisesta käytöstä annetussa laissa (552/2019). Kyseistä lakia on moitittu vaikeaksi soveltaa, ja säädöksen suhde muuhun sääntelyyn on ollut epäselvä.⁷⁶

Informaatio-oikeudellinen sääntely luo siis puitteita, jotka on otettava huomioon suunniteltaessa informaatioon kohdistuvaa käsittelyä, kuten informaation tutkimista. Merkitystä on sillä, minkälaisia tietoaaineistoja informaatiotutkimuksessa käsitellään. Lain-säädäntö turvaa myös tiedonsaannin tieteellistä tutkimusta varten. Sääntely luo myös puitteita tutkimustiedon tiedonhallinnalle, joskin sääntelymäärään vaikuttaa se, kenen hallussa tutkimusaineistot ovat tai kuka toimii rekisterinpitäjänä. Jos tutkimusaineistot ovat korkeakoulun hallussa, sovellettavaksi tulevat julkisuuslaki ja tiedonhallintalaki. Jos tutkimusaineistot ovat tutkijan hallussa ilman sidosta korkeakouluun, tiedonhallinnallisesti tutkimusaineistoja käsitellään lähinnä henkilötietojen suojaa koskevan sääntelyn puitteissa otta-malla huomioon hyvän tieteellisen käytännön mukaiset menettelyt.

Lopuksi

Tässä kirjoituksessa on käyty läpi joitakin näkökulmia oikeustieteen ja informaatiotutkimuksen välisiin yhteyksiin. Oikeusinformatiikka luo näille tieteenaloille yhteisen tutkimusalustan, jota on kuitenkin kansallisesti hyödynnetty varsin vähän tutkimuksessa ja ope-tuksessa. Oikeusinformatiikan tutkimuksen ongelmana erityisesti kansallisesti on ollut tieteidenvälisyyden puute, kun tutkimusalan tutkimustoiminta on ollut lähinnä oikeustieteissä tietoteknologian ilmiöitä kuvailevaa kirjoittelua. Kuitenkin informaatiotutkimuksella ja tietojenkäsittelytieteellä olisi paljon annettavaa oikeusinformatii-kan tutkimukselle ja sen monipuolistamiselle.

⁷⁶ PeVL 25/2021 vp, 4.

Niin informaatiotutkimuksen kuin laajemmin tieteellisen tutkimuksen kannalta katsottuna sääntely-ympäristöön sisältyy joitakin merkittäviäkin puutteita tällä hetkellä. Tutkimustiedon hallintaan ei ole varsinaista sääntelyä, elleivät tutkimusaineistot ole yliopiston hallussa, jolloin julkisuuslaki ja tiedonhallintalaki tulevat sovellettavaksi tällaisten tietoaaineistojen hallintaan tietosuojasääntelyn lisäksi. Tiedeyhteisöllä on omat itsesääntelyn tai myötäsääntelyn kautta luomansa normit tieteelliselle tutkimukselle tieteen eettisinä käytäntöinä, mutta niiden juridinen sitovuus on varsin heikko.⁷⁷

Niin viranomaisten tieteellistä tutkimusta varten arkistoidut tietoaaineistot kuin tutkijoiden muodostamat tietoaaineistot ovat hajanaisesti saatavilla. Suomessa ei ole ollut kansallista tietoaaineistojen metadatan hallinnointiin tarkoitettua palvelua, joka ohjaisi tietoaaineistojen lähteille, mutta tutkimustietovarannosta annetun lain (1238/2021) täytäntöönpano tulee jollakin aikavälillä parantamaan jossain määrin tilannetta. Hajanainen tiedonhallinta ja erilaiset käytännöt johtavat osaltaan tehottomaan tutkimustiedon keräämiseen ja hallintaan, vaikka eri tutkimusten rahoittajat vaativatkin muun muassa tutkimustietoaaineistojen avoimuutta rahoitusehdoissaan. Sääntely-ympäristö ei kuitenkaan luo edellytyksiä tällaisille vaatimuksille, eikä infrastruktuurimmekaan tue tällaista rahoittajien tahtotilaa. Oma oikeudellinen kysymyksensä on, millä toimivallalla erityisesti viranomaisina toimivat tutkimusrahoittajat asettavat rahoitusehdoissaan vaatimuksia tutkimustiedon avoimuudesta, jos niitä ei voida järkevällä tavalla toteuttaa lainsäädännön puitteissa.

⁷⁷ Tutkimuseettinen neuvottelukunta 2023.

Lähteet

- Aarnio, Aulis (1989) *Laintulkinnan teoria*. Porvoo–Helsinki–Juva: Werner Söderström Osakeyhtiö.
- Alexopoulos, Charalambos & Shefali Virkar & Michalis Avgerinos Loutsaris & Anna-Sophie Novak & Euripidis Loukis (2020) Analysing legal information requirements for public policy making. *Electronic participation: 12th IFIP WG 8.5 International conference*, ePart 2020, Linköping, Sweden, August 31–September 2, 2020, Proceedings 12, 95–108. Springer International Publishing.
- Andersson, Helena & Anna Nordin (2021) Säker informationshantering i digital miljöer. Teoksessa Cecilia Magnusson Sjöberg (toim.) *Rättsinformatik - Juridiken i det digitala informationssamhället*. Lund: Studentlitteratur AB, 67–161.
- AOA 4461/4/14. *Eduskunnan apulaisoikeusasiamiehen ratkaisu* 9.4.2015.
- AOA EOAK/101/2019. *Eduskunnan apulaisoikeusasiamiehen ratkaisu* 8.6.2020.
- AOK OKV/471/10/2021. *Apulaisoikeuskanslerin ratkaisu* 22.2.2022.
- Bannan, Christine (2021) Legal data access. Teoksessa Daniel Martin Katz & Ron Dolin & Michael J. Bommarito (toim.) *Legal informatics*. Cambridge: Cambridge University Press, 467–480.
- Bates, Marcia J. (1999) The invisible substrate of information science. *Journal of the American society for information science* 50:12, 1043–1050.
- Bing, Jon (1986) A background analysis for information law. Teoksessa Pekka Lehtinen (toim.) *Oikeus ja uudet ongelmat*. Turku: Oikeustieteen ylioppilaiden yhdistys LEX ry, 43–60.
- Borko, Harold (1968) Information science: What is it? *Journal of the American society for information science* 19, 3–5.
- Dolin, Ron (2021a) Technology issues in legal philosophy. Teoksessa Daniel Martin Katz & Ron Dolin & Michael J. Bommarito (toim.) *Legal informatics*. Cambridge: Cambridge University Press, 5–23.
- Dolin, Ron (2021b) XML in Law. Teoksessa Daniel Martin Katz & Ron Dolin & Michael J. Bommarito (toim.) *Legal informatics*. Cambridge: Cambridge University Press, 61–68.
- Dommering, Egbert J. (1991) An introduction to information law work of fact at the crossroads of freedom and protection. Teoksessa Egbert J. Dommering & P. Bernt Hugenholtz (toim.) *Protecting work of fact – Copyright, freedom of expression and information law*. Dordrecht/Boston: Kluwer Law and Taxation Publishers, 1–40.
- Erdelez, Sanda & Sheila O’Hare (1997) Legal informatics: Application of information technology in law. *Annual review of information science and technology* 32, 367–402.
- Euroopan unionin virallinen lehti (2021) Oikaisu Euroopan parlamentin ja neuvoston asetukseen (EU) 2016/679, annettu 27 päivänä huhtikuuta 2016, luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta (yleinen tietosuojasäätös). *Euroopan unionin virallinen lehti* L 74/35 4.3.2021.
- Finto. Suomalainen asiantuntijapalvelu. <https://finto.fi/fi/>

4.3 Oikeustieteen ja informaatiotutkimuksen yhteyksiä etsimässä

- Gasser, Urs (2002) What is information law - and what could it be? Teoksessa Urs Gasser (toim.) *Informationsrecht in „E“-umgebungen - Information law in environments*. Baden-Baden: Nomos Verlagsgesellschaft.
- Helin, Markku (2017) Siviilioikeuden tutkimusmetodit itsenäisessä Suomessa – matka käsitelainopista mihin? *Lakimies* 7–8/2017, 894–907.
- Hinson, Christopher L. (2005) Legal informatics: Opportunities for information science. *Journal of education for library and information science* 46:2, 134–153.
- Hirvonen, Ari (2011) *Mitkä metodit? Opas oikeustieteen metodologiaan*. Yleisen oikeustieteen julkaisuja 17, Helsinki: Helsingin yliopisto.
- Karhu, Juha (2005) Tilannekohtainen oikeudellinen harkinta ja oikeuslähdeoppi. Teoksessa Jyrki Tala & Kauko Wikström (toim.) *Oikeus – Kulttuuria ja teoriaa – Juhlakirja Hannu Tolonen*. Turun yliopiston oikeustieteellisen tiedekunnan julkaisuja, A-juhlajulkaisut nro 16. Turku: Turun yliopiston oikeustieteellinen tiedekunta, 25–38.
- Kivivuori, Antti (1980) Mitä hyvältä lakitekstilä vaaditaan? *Kielikello* 2/1980. <https://www.kielikello.fi/-/mita-hyvalta-lakitekstilta-vaaditaan-KKO-2023:69>.
- KKO 2023:69. *Korkeimman oikeuden ennakkopäätös 4.10.2023 t. 1389*.
- Koivumaa, Ari (1995) Oikeudelliset informaatiolähteet - lakikirjasta tietoverkkoihin. Teoksessa Petri Martikainen (toim.) *Oikeuskirja*. Rovaniemi: Pandecta, 545–578.
- Koulu, Riikka (2018) Digitalisaatio ja algoritmit – oikeustiede hukassa? *Lakimies* 7–8/2018, 840–867.
- Kuopus, Jorma (1988) *Hallinnon lainalaisuus ja automatisoitu verohallinto*. Helsinki: Lakimiesliiton Kustannus.
- Kuronen, Timo (1998) *Tietovarantojen hyödyntäminen ja demokratia*. Sitra 174. Helsinki: Suomen itsenäisyyden juhlarahasto.
- Lainkirjoittajan opas* (2013) Oikeusministeriön julkaisu 37/2013, Selvityksiä ja ohjeita. <http://lainkirjoittaja.finlex.fi/>
- Lauritsen, Marc (2021) Document automation. Teoksessa Daniel Martin Katz & Ron Dolin & Michael J. Bommarito (toim.) *Legal informatics*. Cambridge: Cambridge University Press, 69–83.
- Lipton, Jacqueline (2003) A framework for information law and policy. *Oregon law review*, 2003:82, 695–778.
- Magnusson Sjöberg, Cecilia (2021) Om rättsinformatik. Teoksessa Cecilia Magnusson Sjöberg (toim.) *Rättsinformatik - Juridiken I det digitala informationssamhället*. Lund: Studentlitteratur AB, 23–31.
- Magrath, Paul (2015) Judgments as public information. *Legal information management* 15:3, 189–195.
- Malmgren, Staffan (2021) Rättslig informationsförsörjning online. Teoksessa Cecilia Magnusson Sjöberg (toim.), *Rättsinformatik - Juridiken I det digitala informationssamhället*. Lund: Studentlitteratur AB, 541–584.
- Mattila, Heikki E. S. (2010) Oikeuslingvistiikka ja oikeusvertailu. *Lakimies* 5/2010, 719–730.
- Miettinen, Samuli (2013) Sähköisen aineiston todistusvoimaisuudesta asetuksen 216/2013 näkökulmasta. *Defensor legis* 3/2013, 411–416.
- Miettinen, Tarmo (2001) *Tieteen vapaus – Julkisoikeudellinen tutkimus tieteenharjoittajan itsemääräämisoikeudesta, tieteen itsekontrollista ja yliopiston itsehallinnosta*. Helsinki: Lakimiesliiton kustannus.

- Mäkinen, Ilkka (2010) Informaatiotutkimuksen tie. Teoksessa Sami Serola (toim.), *Ote informaatiosta. Johdatus informaatiotutkimukseen ja interaktiiviseen mediaan.* Helsinki: BTJ Kustannus, 11–74.
- Määttä, Kalle (2024a) Oikeuslähdeoppi lakien tulkinnassa – I Lähtökohtia tarkasteluun. Edilex 13.11.2024. <https://www.edilex.fi/artikkelit/100010>.
- Määttä, Kalle (2024b) Oikeuslähdeoppi lakien tulkinnassa – V Salitut oikeuslähteet. Edilex 20.11.2024. <https://www.edilex.fi/artikkelit/100014>.
- Määttä, Tapio (2015) Yleislain kunnioittamisen periaate lainvalmistelussa ja laintulkinnassa. Teoksessa Anssi Keinänen, Reima Kukkonen, & Mia Kilpeläinen (toim.) *Oikeustieteiden moniottelija – Matti Tolvanen 60 vuotta.* Helsinki: Edita, 309–329.
- Nay, John J. (2023) Law informs code: A legal informatics approach to aligning artificial intelligence with humans. *Northwestern journal of technology and intellectual property* 20:3, 309–392.
- Niemivuo, Matti (2020) *Lainvalmistelu - oikeudelliset perusteet ja käytäntö.* Helsinki: Helsingin seudun kauppakamari.
- Nuotio, Kimmo (2005) Oikeuslähteet, ”supernormistot” ja ratkaisujen perustelu. Teoksessa Jyrki Tala & Kauko Wikström (toim.) *Oikeus – Kulttuuria ja teoriaa – Juhlakirja Hannu Tolonen.* Turun yliopiston oikeustieteellisen tiedekunnan julkaisuja, A-juhlajulkaisu nro 16. Turku: Turun yliopiston oikeustieteellinen tiedekunta, 127–152.
- Oksanen, Arttu & Minna Tamper & Jouni Tuominen & Eetu Mäkelä & Aki Hietanen & Eero Hyvönen (2019) Semantic Finlex: Transforming, publishing, and using Finnish legislation and case law as linked open data on the web. *Knowledge of the law in the big data age* 317, 212–228.
- PeVL 3/2009 vp. Perustuslakivaliokunnan lausunto hallituksen esityksestä 89/2008 vp laiksi väestötietojärjestelmästä ja Väestörekisterikeskuksen varmennepalveluista.
- PeVL 60/2014 vp, 3. Perustuslakivaliokunnan lausunto hallituksen esitystä eduskunnalle 310/2014 vp laiksi lukiolaissa, ammatillisesta peruskoulutuksesta annetussa laissa ja ammatillisesta aikuiskoulutuksesta annetussa laissa tarkoitetun koulutuksen rahoituksesta ja laeiksi lukiolain, ammatillisesta peruskoulutuksesta annetun lain, ammatillisesta aikuiskoulutuksesta annetun lain 18 §:n ja opiskelijavalintarekisteristä, korkeakoulujen valtakunnallisesta tietovarannosta ja ylioppilastutkintorekisteristä annetun lain 7 ja 9 §:n muuttamisesta.
- PeVL 45/2016 vp. Perustuslakivaliokunnan lausunto hallituksen esitystä eduskunnalle 128/2016 vp laiksi ympäristönsuojelulain muuttamisesta.
- PeVL 31/2017 vp. Perustuslakivaliokunnan lausunto hallituksen esitystä eduskunnalle 72/2017 vp laiksi valtakunnallisista opinto- ja tutkintorekistereistä.
- PeVL 4/2021 vp. Perustuslakivaliokunnan lausunto hallituksen esitystä eduskunnalle 212/2020 vp laiksi sosiaali- ja terveydenhuollon asiakastietojen sähköisestä käsittelystä sekä eräiksi siihen liittyviksi laeiksi.
- PeVL 25/2021. Perustuslakivaliokunnan lausunto hallituksen esityksestä eduskunnalle laiksi sosiaali- ja terveystietojen toissijaisesta käytöstä annetun lain 60 §:n muuttamisesta.
- PeVL 38/2021. Perustuslakivaliokunnan lausunto hallituksen esityksestä eduskunnalle laiksi sisäasioiden rahastoista ohjelmakaudella 2021–2027.

4.3 Oikeustieteen ja informaatiotutkimuksen yhteyksiä etsimässä

- PeVM 25/1994 vp. Perustuslakivaliokunnan mietintö hallituksen esityksestä 309/1993 vp perustuslain perusoikeussäännösten muuttamisesta.
- Pihlajamäki, Antti (2004) *Tietojenkäsittelyrauhan rikosoikeudellinen suoja – Datarikoksia koskeva sääntely Suomen rikoslaissa*. Suomalaisen lakimiesyhdistyksen julkaisuja, A-sarja nro 258. Helsinki: Suomalainen lakimiesyhdistys.
- Pöysti, Tuomas (1999) *Tehokkuus, informaatio ja eurooppalainen oikeusalue*. Helsingin yliopiston oikeustieteellisen tiedekunnan julkaisuja. Helsinki: Helsingin yliopisto.
- Pöysti, Tuomas (2006) Oikeudellisen tiedon niukkuus ja henkilötietojen suoja Teoksessa Aulis Aarnio (toim.) *Syntymästä kuolemaan, oikeudesta informaatioon*. Suomalaisen lakimiesyhdistyksen julkaisuja, E-sarja N:o 17. Helsinki: Suomalainen lakimiesyhdistys, 303–328.
- Repo, Aatto J. (1984) *Tiedon arvo - muistiinpanoja I*. Kirjastotiede ja informatiikka 3:3, 51–60.
- Råman, Jari (2006) *Regulating secure software development - Analysing the potential regulatory solutions for the lack of security in software*. Väitöskirja. Rovaniemi: Lapin yliopisto.
- Saarenpää, Ahti (1999) Oikeusinformatiikka. Teoksessa *Encyclopaedia iuridica fennica* 7, *Oikeuden yleiset tieteet*. Helsinki: Suomalainen Lakimiesyhdistys, 713–726.
- Saarenpää, Ahti (2018) Legal informatics: A modern social science and a crucial one. Teoksessa Peter Wahlgren (toim.) *50 years of law and it*, Scandinavian studies in law, 65, 15–38.
- Saarenpää, Ahti & Juhana Riekkinen (2023) Oikeusinformatiikan perusteet. Rovaniemi: Lapin yliopisto.
- Seipel, Peter (2005) Legal informatics broad and narrow. Teoksessa Cecilia Magnusson Sjöberg (toim.) *Legal management of information systems – Incorporating law in e-solutions*. Lund: Studentlitteratur, 17–36.
- Sharma, Sugam & Samia Gamoura & Deva Prasad & Arti Aneja (2021) Emerging legal informatics towards legal innovation: Current status and future challenges and opportunities. *Legal information management*, 21:3–4, 218–235.
- Siltala, Raimo (2003) *Oikeustieteen tieteenteoria*. Suomalaisen Lakimiesyhdistyksen julkaisuja A-sarja nro 234. Helsinki: Suomalainen Lakimiesyhdistys.
- Stock, Wolfgang G. & Mechtild Stock (2013) *Handbook of information science*. Berliini: De Gruyter Saur.
- Tala, Jyrki (2005) *Lakien laadinta ja vaikutukset*. Helsinki: Edita Publishing Oy.
- Trubow, George B. (1985) Information law overview. *The John Marshall law review*, 18:4, 815–828.
- Tuori, Kaarlo (2002) Käsitelainopin itsepuolustus. *Lakimies* 7–8/2002, 1295–1320.
- Tutkimuseettinen neuvottelukunta (2023) *Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa - Tutkimuseettisen neuvottelukunnan HTK-ohje 2023*. Tutkimuseettisen neuvottelukunnan julkaisuja 2/2023. Helsinki: Tutkimuseettinen neuvottelukunta.
- Valtonen, Marjo (2015) Arvonmääritys. Teoksessa Pekka Henttonen (toim.) *Johdatus asiakirjahallinnan tutkimukseen*. Helsinki: Avain, 93–126.

- Virkar, Shefali & Charalampos Alexopoulos & Sofia Tsekeridou & Anna-Sophie Novak (2022) A user-centred analysis of decision support requirements in legal informatics. *Government information quarterly* 39:3, 101713.
- Voutilainen, Tomi (2009) *ICT-oikeus sähköisessä hallinnossa - ICT-oikeudelliset periaatteet ja sähköinen hallintomenettely*. Helsinki: Edita Publishing Oy.
- Voutilainen, Tomi (2015) *Julkisuus ja henkilötietojen suoja oikeustieteellisessä tutkimuksessa*. Edilex 2015/1.
- Voutilainen, Tomi (2019) *Oikeus tietoon - Informaatio-oikeuden perusteet*. Helsinki: Edita Publishing Oy.
- Voutilainen, Tomi (2021) *Hallituksen asiakirjat - onko niitä?* Edilex 2021/52
- Voutilainen, Tomi & Matti Muukkonen (2021) *Sosiaali- ja terveydenhuollon tiedonhallinnan sääntelyn systematiikkaongelmat*. Edilex 2021/31.
- Voutilainen, Tomi & Juuso Ouli (2023) Virkavelvollisuuksien ja virkatoimien todennettavuuden vaatimus, *Defensor legis* 1/2023, 52–72.
- Wikipedia (2024a) *Legal informatics*. https://en.wikipedia.org/wiki/Legal_informatics
- Wikipedia (2024b) *Oikeusinformatiikka*. <https://fi.wikipedia.org/wiki/Oikeusinformatiikka>